**Analysis of KC-46 Live-Fire Risk Mitigation
Program Testing**

THESIS

Chad Nile Chamberlain, Captain, USAF

AFIT/OR-MS/ENS/12-06

**DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY**

# *AIR FORCE INSTITUTE OF TECHNOLOGY*

**Wright-Patterson Air Force Base, Ohio**

AFIT/OR-MS/ENS/12-06

**ANALYSIS OF KC-46 LIVE-FIRE RISK MITIGATION
PROGRAM TESTING**

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

C. Nile Chamberlain, BS

Captain, USAF

March 2012

**ANALYSIS OF KC-46 LIVE-FIRE RISK MITIGATION
PROGRAM TESTING**

C. Nile Chamberlain, BS
Captain, USAF

Approved:

| | |
|---|---|
| _____//signed//_____ | ___15 March 2012___ |
| Dr. Raymond R. Hill Jr. (Advisor) | date |
| | |
| _____//signed//_____ | ___15 March 2012___ |
| Dr. Darryl K. Ahner (Reader) | date |

AFIT/OR-MS/ENS/12-06

# Abstract

Increased emphasis to include statistical rigor in all testing from the Director of Operational Test and Evaluation (DOT&E) over the past few years has brought an augmented look at testing across the Department of Defense. This work looks at the methodology currently used in live fire testing, particularly involving the risk mitigation of the KC-46 dry-bay test program. It addresses gaps within the methodology in designing as well as analyzing the results of a statistically rigorous test. In addition this research furthers recent work of modeling the characterization of ballistic impact flash by validating concurrent models and characterizing the error due to these models as a function of time and input factors in an attempt to identify systemic bias that may be correctable.

**Dedication**

*To my wife and my children whose unselfish actions allowed me the time and the fortitude to complete this work and inspired me to always come home!*

## Acknowledgments

# Table of Contents

# List of Figures

# List of Tables

# List of Equations

**ANALYSIS OF KC-46 LIVE-FIRE RISK MITIGATION
PROGRAM TESTING**

## 1. Introduction

### 1.1 Background

Design of Experiments has been a form of statistically rigorous testing long used

in the fields of agriculture, science, and industry to maximize the information yielded by

a limited number of test points. Recent guidelines from the Director of Operation Test &

Evaluation have mandated the "increases use of scientific and statistical methods in

developing rigorous, defensive test plans and in evaluating their results" (Office of the

Secretary, 2010). As systems progress technologically their complexity grows nearly

exponentially as do the parameters tested in order to provide indication of the system's

preparedness to continue along the Department of Defense (DoD) acquisitional process.

To test and analyze such systems requires strong statistical understanding.

The push to integrate developmental, live fire and operational testing while

fielding new capabilities quickly has created a void to executing tests with statistical rigor

adequately and correctly. Pressure to include new concept to testers unfamiliar with the

discipline of statistical rigor can result in the incorrect test designs and data analysis. An

effort to field the knowledge across the DoD with relatively limited experts is leaving

many on a academic island with respect to understanding and implementing the

theoretical details of such rigorous statistical planning. Often inference or mention of

DoE is done to meet the requirement with no real theoretical execution to build designs with which to capture the data corresponding to investigating the objectives of the program or system under test.

## 1.2    Problem Statement

This research examines the methodology for designing and executing a live fire risk mitigation test capable of providing statistical and practical evidence between differing armor piercing incendiary projectiles influence upon incendiary flash function and the probability of penetration.  This work also does initial validation of existent and growing models for characterizing the flash function of a projectile against an aircraft.  A key objective of this work is to capture the current methodology used by an agency representative of how live fire is currently planned, executed and analyzed in one sector of the DoD.

Design of experiments, analysis of variance (ANOVA), and logistic regression are used to obtain meaningful results from systematically collected data.  Flash model validation involves visual comparison of actual versus predicted shots as a function of time and a measure of the residual of such plots in an effort to describe the systemic bias that may be correctable within future work to define accurate model coefficients.

An assumption made in this work is that the process studied within the design, execution, and analysis of the FF1 panel testing is representative of the process used throughout the DoD and that the current methods of analysis are fairly consistent and fail to reach the full potential available.

## 2. Literature Review

This chapter addresses the void in methodology documentation currently captured in the design, development and execution of live fire testing and provides a broad view of completed live fire studies. The work makes no attempt to provide an exhaustive summary of all live fire testing from the past, but uses recent, pertinent tests to demonstrate possible strengths and weakness in the current methodology. This chapter also looks at the increasing interest in statistically designed testing within the DoD. As a follow on to multiple AFIT theses, much of the literature review from past works regarding the development for probability and characterization modeling of ballistic impacts are applicable to the fire prediction model validation presented in this research. Portions of these past reviews are thus re-introduced to provide a thorough understanding of the characterization problem for this document and the independent validation of newer models. Lastly, this chapter summarizes the basic principles of logistic regression for use with dichotomous responses.

### 2.1 Live Fire

The current process of live fire testing developed from the evaluation of a systems survivability. A system's survivability is determined by its vulnerabilities which are defined as the inability of that system to withstand a direct passive or active strike while performing within the defined operational environment (Ball, 2003). As a system proceeds through the acquisitional process it is subject to multiple analyses to ensure the system's preparedness when it proceeds to the next event or stage of acquisition. This

process is very complex and rigorous ideally culminating in the operational testing which will ultimately determine the systems readiness to proceed to production.

Survivability testing is an integral portion of DoD aircraft evaluation and developmental testing.  Additional emphasis upon system survivability arose to curtail the rising trend in aircraft losses following the conclusion of each of WWII, the Vietnam War and subsequent conflicts (Ball, 2003).  In March of 1984 aircraft survivability fully incorporated Joint Live Fire testing (JLF) to fill the absence of full scale vulnerability and lethality test data that existed within system survivability evaluation efforts for fielded systems.  JLF testing is funded by the Office of the Secretary of Defense (OSD).  Congress passed the statutory requirement for Live Fire Test and Evaluation of all armored vehicles in 1985 following the Army's Live Fire Test of a Bradley Fighting Vehicle.  This 1985 event led many to conclude that the survivability of equipment and personnel was not being adequately tested (National Research Council, 1995).  The statutory requirement for testing armored vehicles was expanded to all major manned platforms in 1986 by the congressional statutory for Survivability and Lethality LFT&E (Tonnessen, 2011).

The live fire test law (10 U.S.C 2366,1986) requires the testing of a system within the environment and at a threat level most likely to occur while the system is performing the anticipated combat operations.  Live fire includes firing munitions deemed the most likely threat to the operation of a system to determine the vulnerability and susceptibility of a system and its user to attack and the effects upon the system regarding combat performance.  The live fire law (LFL) encourages using a full-up, operationally ready

representative system first at a sub-scale level and later at the full-scale level (National Research Council, 1993).

The definition of full-up and full-scale testing is subject to interpretation by the test committee overseeing the system live fire test (LFT). Most committees agree to define these terms as:

- Full-up testing is defined as a complete or partial system with a full complement of fuel, ammunition, and hydraulic fluids such as will be carried by the system into a combat situation.

- Full-scale testing is defined as testing conducted on a complete or total system that may or may not be full-up representation of the final operational system (National Research Council, 1993).

The objective of live fire testing is to identify any inherent system design weaknesses early enough in the program's acquisition to allow corrective actions to mitigate or eliminate the discovered weakness thus increasing the survivability of the system and its user in combat (National Research Council, 1995). Surprisingly, there are not many documented cases in published venues detailing the planning and conduct of live-fire testing, likely due to little motivation by the tester to publish their experience little requirement to produce a full final live-fire test reports, and issues due to classification of system proprietary information.

### 2.1.1   C-5 Live Fire

As part of the decade long modernization program for the C-5, the Office of the Secretary of Defense (OSD) determined that the C-5 was a system covered under live-fire

test and evaluation. The live fire testing conducted to assess dry-bay fire potential was part of this coverage. Kemp and Woods (2011) primary objective in testing was to determining the probability of dry-bay fire associated with C-5 leading and trailing wing edges as a result of ballistic impact. Along with dry-bay fire probability, the live fire testing of the C-5 legacy wing examined the effectiveness of the fire suppression system in the wing's leading edge and the ballistic damage possibly incurred for three nested hydraulic lines. Testing utilized an outboard, left-hand, wing section from a retired C-5 asset acquired from the Aerospace Maintenance and Regeneration Group (309 AMARG) at Davis-Monthan AFB, Arizona. A rebuild of the internal components and systems brought the section to a full-up configuration. Accurate flying conditions were then created at the 46th Test Group Aerospace Vehicle Survivability Facility (AVSF) at Wright Patterson AFB OH, range 3. Figure 1 shows the section of the wing tested between the dashed lines and its relation to the layout of fuel tanks within the wing structure
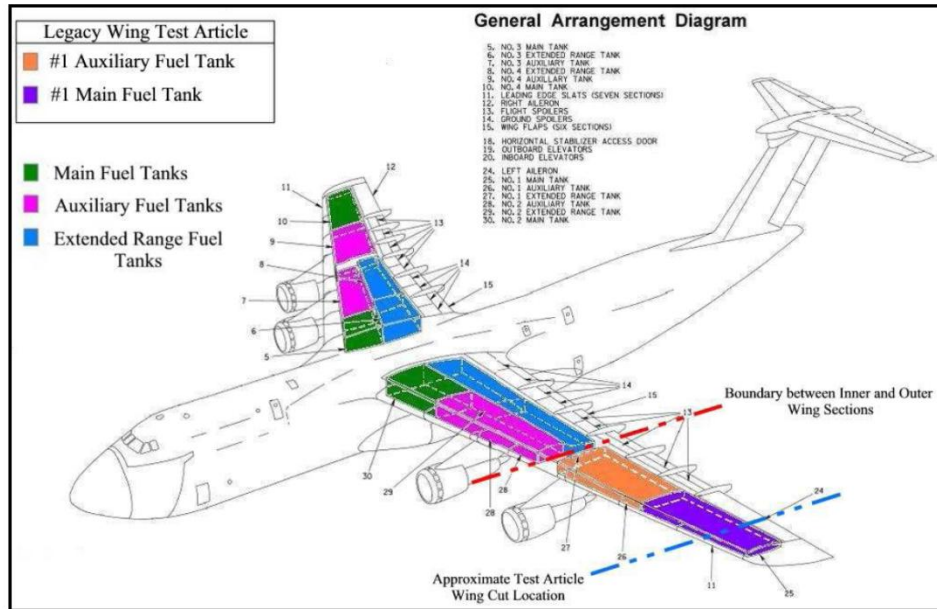
**Figure 1. C-5 Legacy Wing Fuel Layout (Kemp & Woods, 2011)**

The live fire test involved eight shots with two pretests to mitigate the risk associated with regular testing as well as the overall program. Six control factors were considered with three response variables measured for each of the eight shots. Table 1 shows the matrix of the test factors and responses. Program sensitive information was omitted describing the threat type and azimuth angles.

**Table 1. C-5 Legacy Test Matrix and Results**

**Conditions and Fire Results Table from Legacy Wing Testing**

**Summary Results Table**

| Test Event | Threat | Azimuth (deg.) | Elevation (deg.) | Threat Speed (ft/s) | Slats/Flaps | Airflow (knots) | Fuel Level % | Target Bay | Target | Temperature (deg. F) | Fire Type | Fire Duration (sec) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | - | 0 | 85 | 1910 | Retracted | 250 | 0 | Leading Edge | Hydraulic Return Line | Ambient | None | N/A |
| 1B | - | 0 | 85 | 1876 | Retracted | 272 | 0 | Leading Edge | Hydraulic Return Line | Ambient | | N/A |
| 2 | - | 0 | 17 | 2176 | Retracted | 275 | 100 | Leading Edge | Front Spar Web | >1000 | Sustained | +17 |
| 3 | - | 0 | 20 | 1853 | Extended | 161 | 100 | Leading Edge | Front Spar Web | 980 | Self Extinguishing | +14 |
| 4 | - | 0 | 35 | 1542 | Extended | 184.5 | 100 | Trailing Edge | Rear Spar Web | Ambient | None | N/A |
| 4B | - | 0 | 28 | 2013 | Extended | 181 | 100 | Trailing Edge | Rear Spar Web | >1800 | Sustained | +15 |
| 6 | - | 0 | 30.4 | 1990 | Extended | None | 100 (water) | Trailing Edge | Multiple Hydraulic Lines | N/A | N/A | N/A |
| 6B | - | 0 | 44.9 | 2126 | Extended | None | 100 (water) | Trailing Edge | Multiple Hydraulic Lines | N/A | N/A | N/A |

The methods used in the data analysis are not discussed nor was the developmental or consulted expert reasoning behind why the factors and responses were chosen for the test. Alterations were made in the midst of testing, for instance removing a planned run (#5) citing the three previous resulting fires and funding/schedule as rationale. No justification for the removal of this planned shot was provided. The results of the testing while admittedly not producing a solid statistical foundation, were stated to provide a "snapshot from which to draw conclusions based on a solid foundation of experience of the integrated test team" (Kemp & Woods, 2011). The study determined that an incendiary projectile passing through the spar of either the leading or trailing wing edge has a very high probability of resulting in a fire, although that probability was never quantified. It was also found that the fire suppression system did not mitigate nor suppress leading edge dry-bay fires. With a lack of retrievable test methodology it is

nearly impossible to use a sequential design to further analyze the findings of this test or to infer the collected data to a similar objective.

### 2.1.2    F-35 Live Fire Test 2010

Conducted under the direction of Lockheed Martin, test series designator XG-SV-LF-07C (LF-07C) was run to examine the response of the F-35 JSF aircraft and its pilot to a series of system failures representative of damage due to ballistic impact. This test series took place at the Vehicle Systems Processing/Flight Control System Integration Facitiliy (VIF) and the Vehicle Systems Integration Facility (VSIF) at Ft. Worth, TX. All testing was done through the integration of simulation models to represent the result of ballistic interrogations. VSIF resources were utilized minimally for only those test runs requiring the use of real hardware such as electrohydraulic actuators, electrical units, and converter regulators (Andrus, 2010).

The only feasible method for evaluating the objectives without actually shooting a flying aircraft was to use a simulated system with a pilot-in the loop. The test was stated to have improved efficiency of Live Fire Test and Evaluation program but really provides no quantifiable evidence to support the claim. The methodology behind the development of the criterion driving the test was made clear pointing out the highly integrated subsystems critical for aircraft performance and pilot survivability.

LF-07C testing attempted to fill in the gaps left by previous JSF testing regarding the ability of a pilot to accurately and quickly assess the aircraft's remaining capability after sustaining an impact. These main assessments include the aircraft's ability to maintain controlled flight, the time before control is lost, and to determine if the

9

controlled time remaining was sufficient to either get home, fulfill the mission or both (Andrus, 2010).

The Live Fire Team developed a list of test cases to be address based upon identified issues from the JSF Live Fire Test and Evaluation Master Plan (LF-TEMP). Each case within the test matrix had an individual objective. The amount of test runs needed was reduced by cross-examining the LF-07C test matrix with the failure mode and effect testing results to ensure redundant testing was not executed. A total of 40 test cases were built and executed, 31 common to all JSF platforms and nine unique to the short take-off vertical landing (STOVL) variant. Each case was evaluated against three nominally different initial flight conditions with two iterations each. Table 2 shows the three initial conditions. The 31 common tests were evaluated on five criteria and the nine STOVL variant runs against five applicable criteria of their own.

**Table 2.  LF-07C Three Initial Flight Conditions**

| Initial Flight Conditions for LF-07C |
| --- |
| ● 20,000 ft, M0.8, straight and level flight |
| ● 30°dive from 18,000 ft, M0.7 with 4-G pull up to 15° (minimum altitude of about 2,000 ft @ M0.92) |
| ● 20,000 ft, M0.8, 4-G wind-up turn |

Data collection consisted of manual recordings of visual observations as well as digital recordings as functions of time. The digital recordings included the pilot's heads-up display (HUD), Left Multi-Functional Display (MFD), and a screen set to capture pre-determined graphical parameters of pertinent information.

The results were compared to pre-test predictions with 65% matching predictions, 27.5% exceeding predictions and 7.5% below predictions. The pre-test prediction

methodology was not provided.  Further analysis was performed on the non-matching

runs.  Comparison methods or policies were not defined nor were additional analysis

criteria. LF-07C results were used to refine the test matrix for upcoming full-up system

level testing of an F-35 aircraft indicating a sequential testing structure to the JSF Live

Fire Test and Evaluation Master Plan or at a minimum to the LF-07C and the test

proceeding and following it.

No specific details regarding the statistically rigorous planning of the LF-07C

test matrix were provided, neither were the statistical methods used in analyzing the

results of the test though conclusions can be drawn that these principles were part of the

development of the LF-07C.

### 2.1.3   Experimental Design

The two test cases represent too small a sample for meaningful analysis.

However, discussions with various experts confirmed the findings.  In general live fire

testing events are not designed using statistical design consideration, results are often left

unquantified, and planned shots can be changed by personnel running the test.  While

these findings may not be a concern, the question remains whether live fire test programs

might become more effective if statistical design and analysis methods were to be

incorporated.

## 2.2    Experimental Design

Experimental design is the planned and measured alteration of variable inputs to a

system response(s) of interest in an effort to determine the effect of the input variable(s)

upon the system outcome(s).  The growth of modern statistical experimental design over

the past century has culminated in it being among "the most useful, powerful, and widely used applicable statistical methods (Johnson, Hutto, Simpson, & Montgomery, 2012).

The examination of an experimental design involves an analysis of variance to test the equality of several effect means and is a most useful technique in the field of statistical inference (Montgomery, 2008). Two theoretical models are used in such inference, the means model and the effects model. Equation 1 and Equation 2 show the basic form of the means and effects model, respectively. The means model is

$$y_{ij} = \mu_i + \varepsilon_i \begin{cases} i = 1,2, \dots, a \\ j = 1, 2, \dots, n \end{cases}$$ (1)

while defining $\mu_i = \mu + \tau_i$  $i = 1, 2, \dots , a$ produces the effects model.

$$y_{ij} = \mu_i + \tau_i + \varepsilon_{ij} \begin{cases} i = 1,2, \dots, a \\ j = 1, 2, \dots, n \end{cases}$$ (2)

In the effects model $\mu$ is the overall common mean, $\tau_i$ is unique to the $i$th treatment and is called the treatment effect. Using the analysis of variance to test the equality of the treatment means the user assumes the errors of the model themselves are normally and independently distributed random variables with a mean of zero and variance of $\sigma^2$, implying that $y_{ij} \sim N(\mu + \tau_i, \sigma^2)$ (Montgomery, 2008). When all factor levels of the model are fixed or chosen by the experimenter the model is a fixed effects model.  A graphical representation of the effects model is shown in Figure 2.
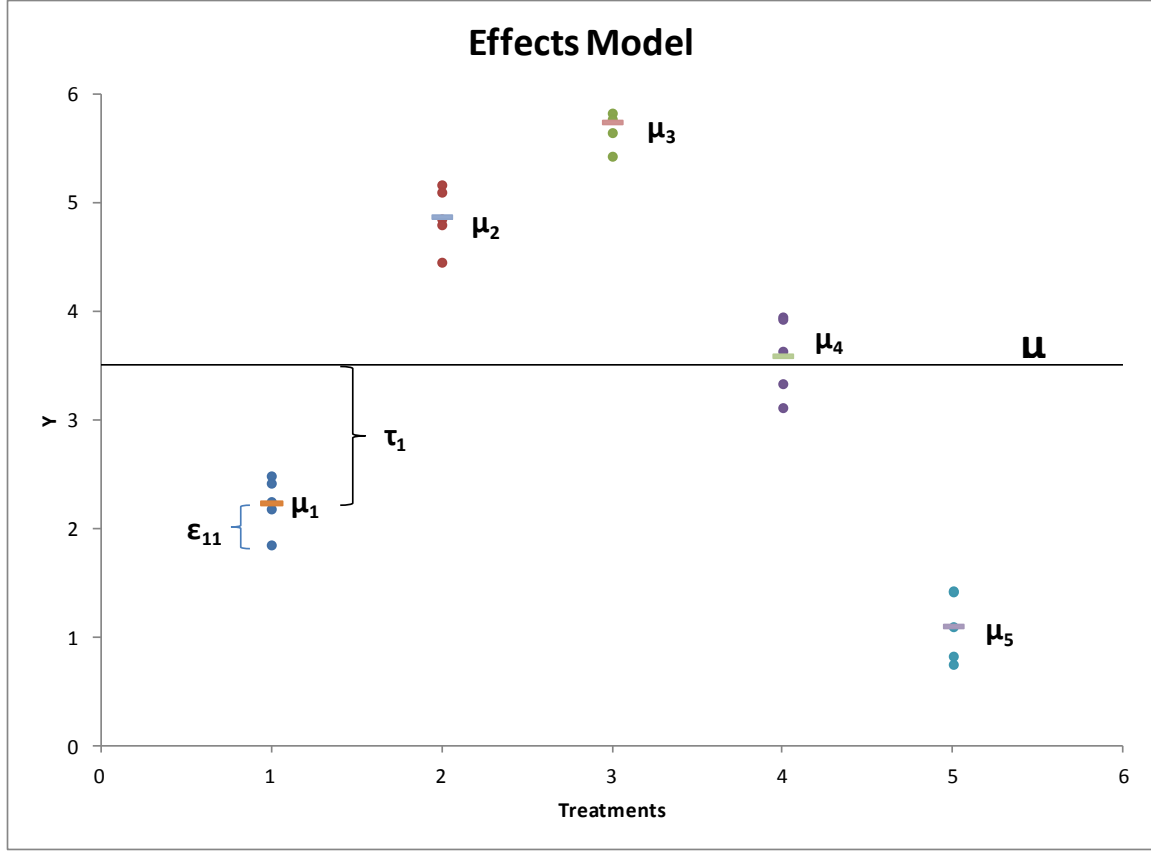
**Figure 2. Graphical Representation of an Effects Model**

With this model the experimenter is testing the equality of the treatments means such to say $E(y_{ij}) = \mu + \tau_i = \mu_i,$ for $i = 1, \ldots,$ k. The corresponding null hypothesis is

$$H_o: \tau_1 = \tau_2 = \ldots \tau_k = 0$$

indicating the treatment levels have no effect on the response variable. The null hypothesis is: $H_a: \tau_i \neq 0 \text{ for at leat one } i, \ i = 1, 2, \ldots, k,$

meaning that at least one treatment has an effect upon the response (Montgomery, 2008).

The ANOVA partitions the total variability in the observations to that associated with each respective treatment, that due the mean, and that due to error. This total sum of squares is usually corrected for the mean and used as a measure of the variability found in the data and is given by:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 \tag{3}$$

The total sum of squares can be partitioned into various components defining $SS_T$ as the sum of the treatment effects and the error. $SS_T = SS_{Treatments} + SS_E$, with $SS_{Treatments}$ being the error between treatment means and $SS_{Error}$ the error within treatment levels.

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$$

$$= n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2 \tag{4}$$

Dividing each of these components by their respective degrees of freedoms produces the mean squares of each term which under a true $H_o$ estimates the error in the model, $\sigma^2$, and is distributed as chi-squared random variables according to Cochran's Theorem (Montgomery, 2008). Dividing $MS_{Treatments}$ by $MS_{Error}$, two independent chi-square distributions, yields an F distributed variable with $a - 1$ and $N - a$ degrees of freedom, where a is the number of treatments and N is the total number of observations, also under a true $H_0$. This variable $MS_{Treatments}/MS_{Error}$ can be compared to an F-statistic, $F_0$, with the same degrees of freedom to indicate the significance of the variance within that treatment with respect to the outcome of the response variable. If $MS_{Treatments}/MS_{Error}$ is greater than the calculated $F_o$ then evidence supports a conclusion that $H_o$ is false and there is at least one $\tau_i \neq 0$, and the statistic F does not follow the F distribution.

The derivation of the fixed effects model above and the estimate of the corresponding values are for a single factor analysis of variance. This derivation is easily expanded to multiple factors.

Design of Experiments (DoE) has long been a standard methodology for testing within the industrial world, proving the impact proper design can have upon the successful reduction of operating costs, increasing outputs, or to simply explore the unknown space defined by an operating environment. Often the implementation of DoE fails because of a gap between those that plan the design and those that execute the design (Coleman & Montgomery, 1993). Coleman and Montgomery building on Hahn (1977) laid out a methodology for designing and executing an industrial experimentation to ensure statistical rigor across all aspects of a design with an industrial application.

Currently experimental design is part of a large scale effort to replace budget and schedule driven testing that has long been the established norm within the DoD testing community. Over the past decade, leadership in DoD has seen statistical designed experiments as a viable way to extract meaningful data from a system test limited in budget and/or resources as the current national budget is focused to reduced spending across the DoD (Johnson, Hutto, Simpson, & Montgomery, 2012).

Previous works performed by students at AFIT have shown the research benefits of statistically rigorous testing. Tallafuse's (2011) work particularly shows the benefit of test planning involving the principles of DoE. His work is detailed later in this chapter.

## 2.3 Flash Characterization

With the majority of Air Force systems subject to the live fire law being aircraft or their subsystems, flash characterization is critical to determining the survivability of a system or subsystem. As a projectile contacts the surface of the airframe structure many factors play a role in determining the dichotomous response of a flash or no flash. The pursuant characteristics of a resulting flash drives the probabilities of the impact resulting in either an un-sustained fire, a sustained fire resulting in a relative degradation of system performance or in the failure of the system. The importance of determining the characterization of a resultant flash and the variables that affect the likelihood of its occurrence have driven research regarding this topic over the past decades as systems become more complex, threats increase and costs rise.

### 2.3.1 Incendiary Function Probability

Four AFIT theses from the early 90s evaluated the probability of flash occurrences and the probability of projectile penetration. Incendiary functioning is defined as the presence of material oxidation due to the residual kinetic energy of a projectile impacting airframe material resulting in a flash or function. Before this research, the prediction of incendiary function drew upon the Penetration Equations Handbook for Kinetic-Energy Penetrators, published by the Joint coordinating Group for Munitions Effectiveness (JTCG/ME). This handbook characterized the prediction of incendiary functions based on a specific target material and separated functioning of a projectile into five categories. Any material or projectile type note specified within the

handbook required correction factors within the determining equations and were not accurate (Talafuse, 2011).

Reynolds (1991) used multivariate analysis and response surface methodology to draw conclusions regarding the incendiary functioning of armor-piercing incendiary (API) projectiles impacting composite material. Reynolds developed two regression models; the first determined an entry or front face function capable of igniting fuel and the second classifying the event as a non-function. Reynolds analysis had four input variables: impact velocity, impact obliquity angle, impact mass and material thickness with three measured responses: residual mass, residual velocity and incendiary function. Reynolds' work expanded the accurateness of the JTCG/ME but did not do well defining the classification of the functioning that was predicted to occur.

Knight (1992) used Reynolds' work to improve the prediction of residual velocity and mass of the projectile as well as the prediction of incendiary functioning. Lanning (1993) furthered the classification of functioning by examining penetration probabilities of a projectile using neural networks and discriminate analysis. The bulk of Lanning's work concluded that composite materials required higher velocities to function but produced longer lasting flashes when compared to aluminum. Blythe (1993) attempted to establish a methodology to build a characterization model for exit side ballistic flash but was only able to recommend a focused velocity regime for composite materials and stated that discriminate analysis would be best for developing a prediction model.

### 2.3.2  *Incendiary Flash Characterization*

Recent technological advances in high speed video have allowed ballistic impact flashes to be captured and mathematically analyzed producing more reliable and reproducible data.  These advancements were used to model the characterization of ballistic flashes (Bestard & Kocher, 2010). Bestard and Kocher's methodology used image processing algorithms.  A data analysis tool was developed to achieve uniform data reduction increasing the accuracy and validity of any subsequent models developed from the test data.  This tool used image processing algorithms to analyze the digital video frame-by-frame and enclose the defined function in an ellipse using least squares minimization.  The analysis showed that the various characteristics of a flash function could be quantitatively described and that clear patterns existed for function position and size.  It was found that the flash position followed a logarithmic trajectory of the projectile's path and that flash size exhibited a Weibull shaped distribution over time.  Orientation of the flash cloud showed no clear trend and was defined as the average of the orientation time series.

Bestard and Kocher conjectured that a complete ballistic impact flash characterization model capable of predicting flash position, size, orientation and thermal energy released as a function  of time could be developed using this methodology.  Such a model would use projectile properties, target properties, impact conditions and ambient conditions as influence factors and predict the flash over time.

Henninger (2010) built a time-based empirical function to model the flash-event time-series data.  He modeled entry-side (front face) flash using time as the regressor. The designed experiment used to collect the data varied projectile velocity, projectile

weight, target panel thickness, and impact obliquity. Flash position, orientation, duration, or its thermal properties were not considered within the scope of the research. The original focus of Henninger's research was to develop a model in the form

$$FlashRadius = f(time) + N(0, \sigma^2)$$

where f(time) is the regression-based model and $N(0, \sigma^2)$ is the noise or error from a normally distributed system. Initial analysis showed that a quartic model provided an acceptable estimate of the flash radius and was of the form:

$$r_{xi}(t) = \beta_{x1}t + \beta_{x2}t^2 + \beta_{x3}t^3 + \beta_{x4}t^4 + \epsilon_{xi}$$
$$r_{yi}(t) = \beta_{x1}t + \beta_{x2}t^2 + \beta_{x3}t^3 + \beta_{x4}t^4 + \epsilon_{yi}.$$

(5)

Replicate runs of the same design showed an averaging effect and resulted in a decline in model accuracy. Model residuals indicated a non-normal distribution and non-constant variance over time. Henninger concluded a better model for flash radius would include a time based error in the form

$$FlashRadius = f(time) + g(time).$$

Henninger's results, combined with the work of Bestard & Kocher, laid the ground work for more accurate meta-models.

### 2.3.3   Meta-Model Development

Talafuse (2011) used the data from Henninger's research as post-processed using the Bestard and Kocher method to build a model for flash prediction using the independent variables of target panel thickness, obliquity angle, projectile mass and initial velocity. Table 3 shows the settings for each independent variable in the actual experiment.

**Table 3. Talafuse Designed Experiment Factors & Levels**

| Factor | Variable | Low | Medium | High | Units |
|---|---|---|---|---|---|
| Panel Thickness | Thick | 0.1 | 0.25 | 0.35 | Inches |
| Obliquity | Angle | 0 | N/A | 45 | Degrees |
| Projectile Mass | Mass | 40 | N/A | 70 | Grams |
| Initial Velocity | Vel | 4000 | 5500 | 7000 | fps |

Talafuse found limitations in the previously collected data. Several shots were truncated because the flash clouds were not caught entirely within the camera frame or were obscured from view by test equipment. This resulted in only 21 of the original 72 shots being usable for statistical analysis leading to the empirical model. As a result Talafuse designed a full factorial model that was not run in time to be examined in his research but was used in future refinement of his resultant meta-model. This design effort included changes to the test configuration to reduce the percentage of unusable shots

Talafuse pointed out that any "analytical model of a ballistic impact flash event must be a function of the input parameters defining that event" (Talafuse, 2011). Equation 6 predicted the time-based quartic model regression coefficients given the respective factor settings and was the method Talafuse devised to relate the factors to the flash radius. The predictive flash radius model is shown in Equation 7.

$$\beta_i = b_0 + b_1 Thickness + b_2 Angle + b_3\ Mass + b_4 Velocity, \quad i = 1, \dots, 4 \qquad (6)$$

$$FlashRadius(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 \qquad (7)$$

Peyton (2012) used the data collected based on Talafuse's full factorial experimental design to analyze entry (front face) flash. There were 283 shots suitable for analysis and Peyton determined a Weibull distributed function would provide a better fit for the data verses Talafuse's original quartic model. Peyton's baseline model became:

$$FlashRadius_x(t) = \gamma_x \left( \left(\frac{\alpha}{\beta_x}\right)\left(\frac{t}{\beta_x}\right)^{\alpha-1} e^{-\left(\frac{t}{\beta_x}\right)^{\alpha}} \right)$$
$$FlashRadius_z(t) = \gamma_z \left( \left(\frac{\alpha}{\beta_z}\right)\left(\frac{t}{\beta_z}\right)^{\alpha-1} e^{-\left(\frac{t}{\beta_z}\right)^{\alpha}} \right) \tag{8}$$

Like Talafuse, Peyton generated a meta-mocel to predict flash radius model parameters as a function of fragment impact parameters:

$$Model\ Coefficient = e^{\wedge}(b_0 + b_1(vel) + b_2(obliq) + b_3(thick) + b_4(mass) \tag{9}$$

Peyton partitioned the data into two sets to allow for cross-validation and after successful validation used the entire data set to build a final model to predict entry (front flash).

Koslow (2012) performed concurrent research with Peyton focused on an exit (back face) flash prediction model conditional on the probability of projectile penetration. Koslow used the same methodology as Peyton (Equation 8 and Equation 9). The Peyton and Koslow models were delivered to ASC/EN for incorporation into the survivability tools. Part of this research involves using live fire test results to independently validate these models.

## 2.4    Logistical Regression

When a response variable has a dichotomous outcome, ANOVA becomes inappropriate due to violation of the error assumptions of the linear statistical model.

Once the differences between the dichotomous response and the linear regression, namely the underlying assumptions and choice of parametric model, are addressed the analysis follows the same general principles (Hosmer & Lemeshow, 1989). Dichotomous responses violate many of the assumptions of Ordinary Least Squares (OLS) regression. Key among these assumptions are that of homogeneity of variance and the normality of errors. OLS regression also produces a model whose prediction range falls between negative infinity and infinity. This does not adequately fit a dichotomous response whose value indicates the presence or absence of an event (Menard, 2002). This lack of appropriateness is apparent when looking at a plot of a dichotomous response with respect to the input factors as shown in Figure 3. No linear model fitted to the data could provide an accurate prediction of the response. A logistical regression of this same data produced a model with a continuous response range from 0 to 1 indicating the probability of a response of "1" given the value of the input factor (Menard, 2002).
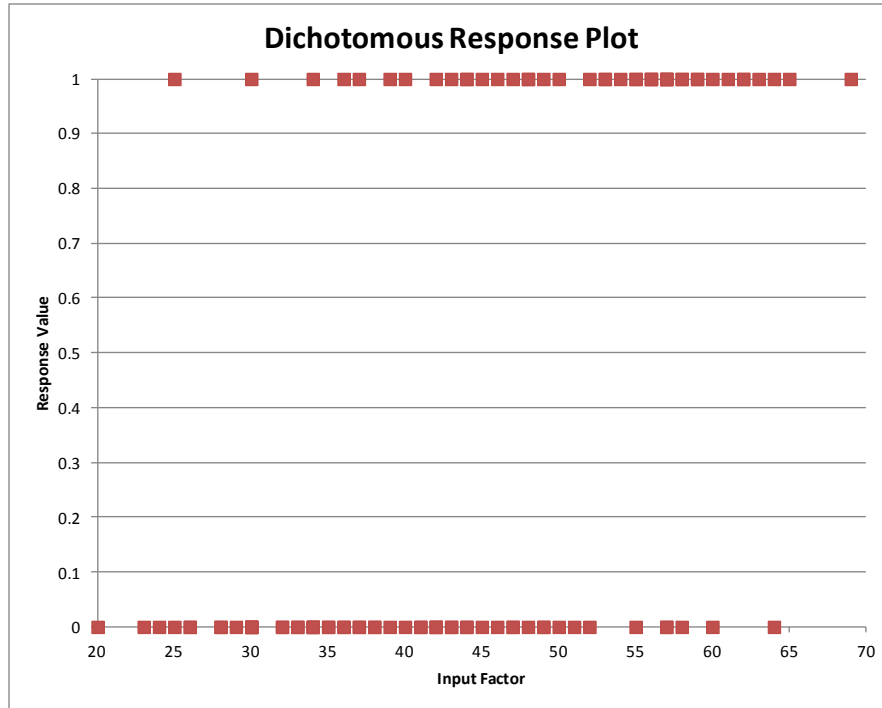
**Figure 3. Plot of a Dichotomous Response**

Figure 4 shows the data from Figure 3 where the input factors have been

partitioned into 8 groups of approximately equal size and the percentage of the responses

equal to a value of "1" within that group plotted against the midpoint of that group.
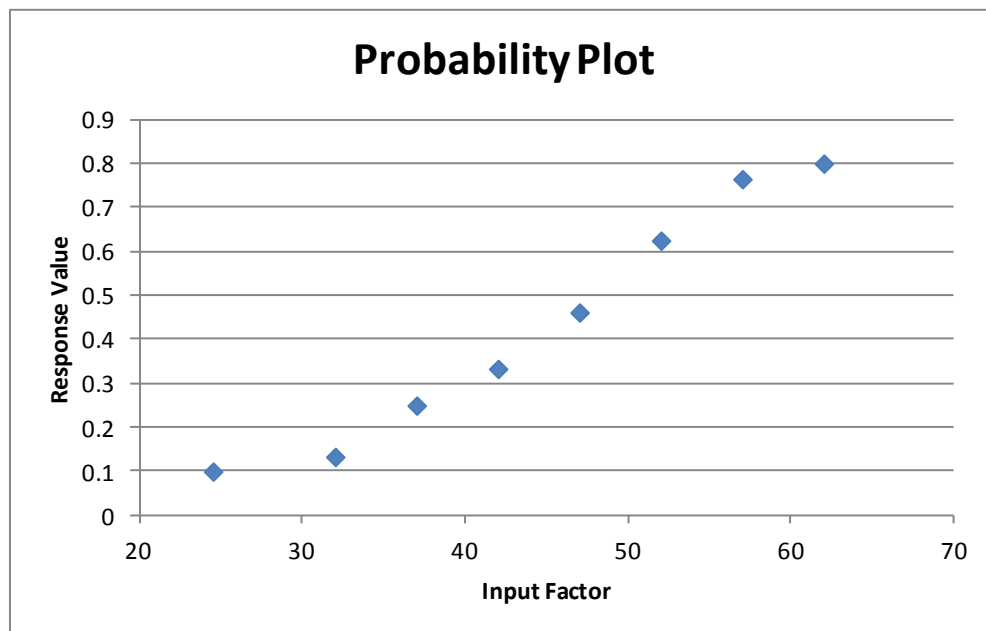


**Figure 4. Probability of Response Plot**

23

Logistic regression determines the probability of the response y equaling "1"

given the input factor $x_i$ and can be expanded to include $x_i$ as a vector.

$$E(Y|x) \; = \; P(y = 1|x_i) \; or \; P(y = 1|\vec{x_i})$$

Letting $\pi(x) = E(Y|x)$ the logistic distribution used to model the probability takes the

form:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \tag{10}$$

The logit transform of this function is important because its characteristics are those of a

linear regression model, being linear in its parameters and allowing for a response range

from negative infinity to infinity based upon the range of the input factors (Hosmer &

Lemeshow, 1989). The transform yields the corresponding coefficient values and is

expressed as:

$$g(x) = \ln\left[\frac{\pi(x)}{1 + \pi(x)}\right] = \beta_0 + \beta_1 x \tag{11}$$

As with the derivation of the ANOVA in section 2.2, the single factor model of the

logistic model and the logit transform easily extend to multiple factors. The assumed

outcome for the observed model, $y = E(Y|x) + \varepsilon$, does not have an error component that

follows a normal distribution with constant variance due to the logistical transforms.

Rather the error component, $\varepsilon$, can take on two values. When $y = 1$ then $\varepsilon = 1 - \pi(x)$ and

when $y = 0$, $\varepsilon = -\pi(x)$ with probability $\pi(x)[1 - \pi(x)]$. The conditional distribution of the

outcome is a binomial distributed variable whose probability is given by $E(Y|x)$ or $\pi(x)$

(Hosmer & Lemeshow, 1989).

The likelihood for a particular pair of input factor and response variable,$(x_i, y_i)$, is expressed in the likelihood function.

$$\zeta(x_i) = \pi(x_i)^{y_i}[1 - \pi(x_i)]^{1-y_i} \tag{12}$$

The overall likelihood function for all pairs, $(x_i, y_i)$ is the product of all $\xi_i$ because the observations are assumed to be independent. This produces a general likelihood function of the form:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} \zeta(x_i) \tag{13}$$

The computation of the value of $\boldsymbol{\beta}$ is done iteratively to calculate the respective values which will maximize the likelihood function (Hosmer & Lemeshow, 1989). The power in a logistical regression is that given a probability of occurrence ($\pi(x)$) the odds of the occurrence is given by $e^{g(x)}$ or $e^{(\beta_0+\beta_1 x)}$.

$$e^{g(x)} = e^{\ln\left[\frac{\pi(x)}{1+\pi(x)}\right]} = \frac{\pi(x)}{1 + \pi(x)} \; or \; e^{\beta_0+\beta_1 x} \tag{14}$$

The ratio of odds for a one unit increase in x is then simply the exponential of the corresponding coefficient or $e^{\beta_i}$ (Wolf, 2012).

## 3. Methodology

This chapter captures some of the methodology of test design within the 46th Test Group to include collaboration on data collection and reduction with Skyward Ltd and InDyne Inc. Section 3.1 discusses this collaborative design development. Section 3.2 illustrates the live fire test execution and the encapsulated data collection process. Section 3.3 explains the analysis methods used to mine and reduce the data from raw range data to usable statistically valued responses. Section 3.4 examines the validation methods used to validate the flash characterization boundary model built by Peyton (2012) and Koslow (2012) which built upon the initial research by Talafuse (2011) using Bestard and Kocher (2010) methodology.

### 3.1     Design of Live Fire Test

As an acquisition category I (ACAT I) program the KC-46 must undergo live fire testing as part of developmental testing. The results of live fire test feed the specification compliance efforts for the airframe as well as the Vulnerability Analysis Report (CDRL A0009). The results are delivered to OSD/DOT&E/LFT&E upon completion and are incorporated into the overall LFT&E Consolidated Final Report due to Congress 90 days prior to the full-rate production decision.

The FF1 panel live fire test was developed as part of the risk reduction of the dry-bay testing for the production KC-46 article. The designator FF1 refers to the first test matrix of the live fire risk reduction testing. FF, or flammable fluid, refers to the purpose of the test; to analyze the flammability of the structure. The objective was to mitigate or reduce the number of shots needed against the production article during the future live

fire testing of the dry-bay.  On a production article large amounts of testing are expensive both in time and budget.  Often limitations of material or personnel availability restrict the number of tests performed.  A risk reduction for the KC-46 Dry Bay production article seeks to describe or define the areas within the experimental space that are fairly stable with respect to specific responses.  Certain shot locations may not be affected by different shot angels, projectile velocities or projectile types and show a resiliency across these factors.  For the specific response of flash function, this translates to either a function or the absence of a function.  By running the panel test as a pre-screening design, analysts determine which areas within the experimental space show larger variance in the response and will therefore need further study during the article testing.

For KC-46 an additional aspect of the test incorporated the fact that within the specification threat of the Alternative Test Plan (ATP) Boeing specified that the characteristics of a s7.62 39mm armor piercing incendiary device be captured.  Limited live fire data was available for the specific API identified and its characteristics were not well defined.  To help develop this characterization, a comparison with a well documented and characterized API, the 7.62 54mm, was made to understand the new threat.  The original purpose to compare various aspects of the two projectiles to include: their physical characteristics, any variance of incendiary within each type, and different burn characters of the incendiaries.   Other aspects of interest were to be accomplished using non-statistical analysis.  The main statistical comparisons were to be done primarily against an aluminum target because of the extensive data for the well tested 7.62 54mm API against the aluminum material.  The statistical comparison of the two API projectiles is the bulk of this current research.

Two portions of the FF1 panel testing matrix are analyzed in this research. The first is the comparison of the two API characteristics against an aluminum target. The seconds is the characterization of varying APIs and steel fragments against a honeycombed composite material. Although the aluminum comparison was part of the entire FF1 panel test matrix, its design and execution were done separate from the other aspects within the test matrix and then concatenated into the FF1 matrix. The remaining runs within the FF1 matrix, beyond the API comparison, were designed as one test split into multiple blocks. Due to customer demands, administrative requests and material constraints this test matrix was not run as designed.

The 7.62 API comparison test investigated the flash characterization of differing armor piercing incendiary (API) projectiles when fired upon production representational composite panels. A $2^4$ full factorial design was initially developed, with the 46th Test Group taking the lead on design considerations and requirements. The discussion paper for the initial designs is in Appendix A. Table 4 shows the factors and their corresponding levels determined by the subject matter experts and influenced by previous testing.

**Table 4. Experiment Input Factors and Levels**

| Factor | Levels | |
| --- | --- | --- |
| | Low | High |
| Threat | 7.62 39mm | 7.62 54mm |
| Velocity | 1500 fps | 2500 fps |
| Obliquity | 0° | 45° |
| Thickness | 0.16 in | 0.25 in |

The input factors for this design were heavily affected by the model formulated by Bestard and Kocher (2010) and used for fire prediction. These previous tests have shown that the four factors of panel thickness, obliquity angle of impact, projectile mass/type, and initial velocity influence any flash upon impact of a fragment/API. Threat levels were included in the place of projectile mass for the API tests since the objective of these tests were to determine the difference, if any, between the 7.62 39mm API and the 7.62 54mm API. The clarification of the additional threat came from the inclusion of this threat within the CDRL A009 for the KC-46. Table 5 shows the initial 7.62 API test matrix in standard order for both natural and coded variables.

**Table 5. Initial Matrix Design in Natural and Coded Variables**

| | Threat | Velocity | Obliquity | Thickness | Threat | Velocity | Obliquity | Thickness |
|---|---|---|---|---|---|---|---|---|
| 1 | 39mm | 1500 | 0 | 0.16 | -1 | -1 | -1 | -1 |
| 2 | 54mm | 1500 | 0 | 0.16 | 1 | -1 | -1 | -1 |
| 3 | 39mm | 2500 | 0 | 0.16 | -1 | 1 | -1 | -1 |
| 4 | 54mm | 2500 | 0 | 0.16 | 1 | 1 | -1 | -1 |
| 5 | 39mm | 1500 | 45 | 0.16 | -1 | -1 | 1 | -1 |
| 6 | 54mm | 1500 | 45 | 0.16 | 1 | -1 | 1 | -1 |
| 7 | 39mm | 2500 | 45 | 0.16 | -1 | 1 | 1 | -1 |
| 8 | 54mm | 2500 | 45 | 0.16 | 1 | 1 | 1 | -1 |
| 9 | 39mm | 1500 | 0 | 0.25 | -1 | -1 | -1 | 1 |
| 10 | 54mm | 1500 | 0 | 0.25 | 1 | -1 | -1 | 1 |
| 11 | 39mm | 2500 | 0 | 0.25 | -1 | 1 | -1 | 1 |
| 12 | 54mm | 2500 | 0 | 0.25 | 1 | 1 | -1 | 1 |
| 13 | 39mm | 1500 | 45 | 0.25 | -1 | -1 | 1 | 1 |
| 14 | 54mm | 1500 | 45 | 0.25 | 1 | -1 | 1 | 1 |
| 15 | 39mm | 2500 | 45 | 0.25 | -1 | 1 | 1 | 1 |
| 16 | 54mm | 2500 | 45 | 0.25 | 1 | 1 | 1 | 1 |
| | | Natural Variables | | | | Coded Variables | | |

Digital high speed cameras captured footage of each shot and with additional numerical analysis of the data yielded five response variables pertaining to the API comparison portion of this research. These responses were continuous measures of front

face flash duration, back face flash duration, and change in projectile mass as well as dichotomous indication of flash occurrence and projectile penetration as well as. Photodiodes gathered additional data needed to obtain flash size measurements on some of the shots.

The additional runs of the FF1 test matrix outside the 7.62 comparison runs used differing projectiles and target material. The design of the entire FF1 test matrix was not included as part of this research but was exploited to obtain additional objectives within this work. The design of the 48 run sub-matrix of the 7.62 comparison test was the only portion of the FF1 test matrix completely within the scope of this effort. The complete FF1 test matrix, including the full 7.62 comparison matrix, is available from AFIT/ENS. Portions of these additional FF1 test points were to validate the Peyton (2012) and Koslow (2012) models and as discussed in Chapter 5. For each subsequent section of this chapter, both the 7.62 API and the remaining runs of the FF1 test are discussed and differentiated where applicable.

## 3.2    Test Execution and Data Collection

The test was conducted by the 46th Test Group, Aerospace Survivability Analysis Branch, at Wright Patterson Air Force Base, OH. While the 46th TG had ultimate control over the test execution, Skyward Ltd and InDyne Inc conducted the physical execution of the individual test points and subsequent data collection and analysis. The current test process follows the guidelines discussed by Coleman & Montgomery (1993) with regard to the execution of a test designed with statistical rigor. Unfortunately, actual adherences to the suggested guidelines for fulfilling and analyzing properly designed test were not

always accomplished.  The testing process was conducted over the course of three months during the end of CY 2011 and beginning of CY 2012 on the range facilities at Wright Patterson Air Force Base.   Execution of the 7.62 API comparisons was accomplished during a three week period near the beginning of the entire test execution.

All projectiles fired were API rounds or steel fragments of varying sizes.  The target panels were roughly eight inches square aluminum panels of varying thickness and temper or composite material representative of material to be used on the external portions of the KC-46 airframe.  The portions of the FF1 matrix used in the 7.62 comparison utilized all API projectiles and aluminum panels of either 0.16 or 0.25 inches thick as indicated in Table 5.

Test set up for all runs were identical for each projectile type.  Break paper was used to calculate the actual projectile velocity at impact.  Placement of this paper at known distances allowed computational verification of projectile velocity just prior to impacting the target as well as the residual velocity after the impact.  Three different events were used to verify the timing of multiple data collections.  The first event is termed an advanced event which occurred 2 feet beyond the target panel.  The second event was the actual panel strike.  The third was the projectile striking a small mound behind the target known as the bullet catcher.  The test set up is illustrated in Figure 5 and Figure 6.  All lighting of the testing was achieved using LED lights allowing a consistent ambient temperature within the range and of the target material.
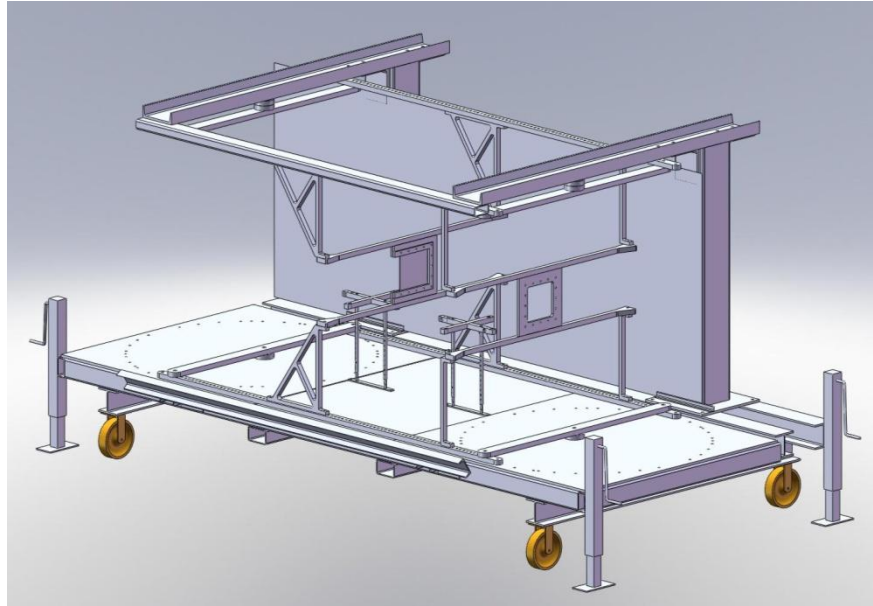
**Figure 5. Depiction of Test Setup**



**Figure 6. Actual Test Set Up**

It is assumed that all tests were conducted with sufficient controls in place to minimize inherent noise within the system and that the randomization of the runs, while

not completely random due to range limitations and demands, was adequate. Persons performing test set-up and execution used the same techniques and methods to mitigate any introduced noise from individuals performing test set-up or execution.

Although the designed portion applicable to this research was agreed upon as the executable experiment, alterations occurred from the time of design to the time of execution. These alterations only applied to the 7.62 API comparison portion of the FF1 matrix. First, the randomized order of the runs was altered yielding a design executed as a split plot design. This was done to accommodate the increased time and difficulty involved with adjusting the obliquity angle. Generally, executing a design differently than planned is ill-advised. Specifically, executing a design erroneously as a split-plot can severely bias results (Cohen, 2010). Fortunately, this adjustment and subsequent run repetitions were done in a way that it did not affect the analysis portion of the test. The second alteration was more serious. Instead of running the test at 0 and 45 degrees for the obliquity angle, it was run at 0 and 60 degrees. This change extended the originally designed space of the model and went against the subject matter experts recommendations for maximum obliquity angle of 45 degrees. A third alteration to the original design varied the replications between all runs. Instead of three replications at each point, replications varied from one to five. This changed the variance characteristics of the design. The predictive variance was no longer constant across the design space but rather became a function of the location within the design.

Just as builders should not deviate from the engineering build plans, test executors should not deviate from the statistically engineered test plan. Such deviations, or alterations, can severely impact test effectiveness and degrade efforts to answer test

objectives. Such alterations appear to be common practice during live fire and thus require increased test execution discipline.

## 3.3    Analysis Method

Two analysis methods were used on the responses collected during the 7.62 API comparison experiment. For continuous responses; Front Face Flash, Back Face Flash, and Panel Weight Change, ANOVA was utilized. For the nominal dichotomous responses involving front face function, back face function and projectile penetration, logistical regression was utilized which produced a probability of occurrence and an odds ratio of unit factor changes upon response occurrence.

### 3.3.1   ANOVA

An analysis of variance (ANOVA) based on an effects model partitions test variance attributed to each (input) factor. If this partitioned variance differs from experimental error, the effect of changing the levels of that factor is deemed significant. Of interest is whether a change in the threat factor setting from "low" to "high" yields a change in the response. For significant factors, effects for each factor level ($\tau_i$) were estimated and individually examined for statistical significance. Significant effects are interpreted as non-zero effects.

### 3.3.2   Logistical Regression

Simple linear regression, and subsequent analysis of variance, of dichotomous responses may produce predicted values that do not lie within the actual range of the dependent responses. Linear regression models assume errors are normally distributed

with constant variance but these assumptions fail to hold for dichotomous response models. For these specific variables it is often more appropriate to predict into which of the two cases the response will fall into based upon the value of the dependant variables or factors. A logistic regression model provides this capability.

Front face flash (FFF) function was the only dichotomous response able to be analyzed due to instability in the models of the other two responses of back face flash function and projectile penetration. The instability was caused by the non-convergence of the estimated coefficients during the iterative calculation of the maximum likelihood function (Equation 13) due to a probability equal to 0 or 1 for the response. A suggested resolution to these unstable response models is discussed in sections 4.1.3 For FFF function, a response value of -1 indicated an absence of function and a value of 1 indicated the presence of flash. The response model was analyzed in JMP using a forward stepwise approach to determine input factors considered as significant. A p-value for a factor to enter the model was 0.05 and a p-value for a factor to leave the model was 0.15. The natural variable settings were used for the continuous input factors to allow the calculation of the odds ratios for one unit increments of the input factors.

Front face flash (FFF) function parameter estimates using the above methodology are shown in Table 6.

**Table 6. Front Face Flash Function Parameter Estimates from Logistic Regression**

**Parameter Estimates and Significance**

| Term | Estimate | Lower 95% | Upper 95% | Std Error | Chi Square | Prob> ChiSq |
|------|----------|-----------|-----------|-----------|------------|-------------|
| Intercept | 1.408933 | -4.208178 | 7.5483201 | 2.908917 | 0.23 | 0.6281 |
| Threat | 1.49406 | 0.5765264 | 2.7206658 | 0.5303128 | 7.94 | 0.0048 |
| Natural Velocity | -0.0004 | -0.002206 | 0.0013026 | 0.0008671 | 0.21 | 0.6437 |
| Natural Obliquity | -0.04966 | -0.089929 | -0.019226 | 0.017521 | 8.03 | 0.0046 |
| Natural Thickness | -4.23126 | -27.37042 | 15.16441 | 10.36086 | 0.17 | 0.6809 |

The unit odds ratios for the factors corresponding to FFF function are given in Table 7.

Note that the odds ratio for the threat is over the entire range of the factor or from -1 to 1,

and corresponds to a two unit increase in threat (Odds Ratio $= e^{(2*\beta 1)}$). This is due to the

categorical classification of threat and the coding of this factor to be -1 for 7.62 39mm

API and 1 for the 7.62 54mm API.

**Table 7. FFF Function Odds Ratio for Unit Increase in Input Factors**

**Unit Odds Ratio**

| Term | Odds Ratio | Lower 95% | Upper 95% |
|------|------------|-----------|-----------|
| Threat | 19.84835 | 3.1678489 | 230.74926 |
| Natural Velocity | 0.999599 | 0.997797 | 1.001303 |
| Natural Obliquity | 1.050913 | 0.913996 | 0.980957 |
| Natural Thickness | 0.014105 | $1.3 \times 10^{-12}$ | 3853183 |

Model variables did not show a linear lack of fit in the logit. The receiver

operating characteristic (ROC) curve and confusion matrix are shown in Figure 7 and

Figure 8, respectively. The ROC curve is a graphical representation of the "true positive"

responses rate verse the "false positive" response rate. A plot to the upper left of the

graph indicates a more accurate model or less "false positives". The confusion matrix is

another way to describe the accurateness of a model with the responses assembled into a

matrix form. The rows represent the true classification of the response and the column the predicted classification. Higher numbers on the diagonal indicate a more accurate model. The area under the curve for the ROC is 0.84725 and the hit rate for the confusion matrix is 0.854 for classification.
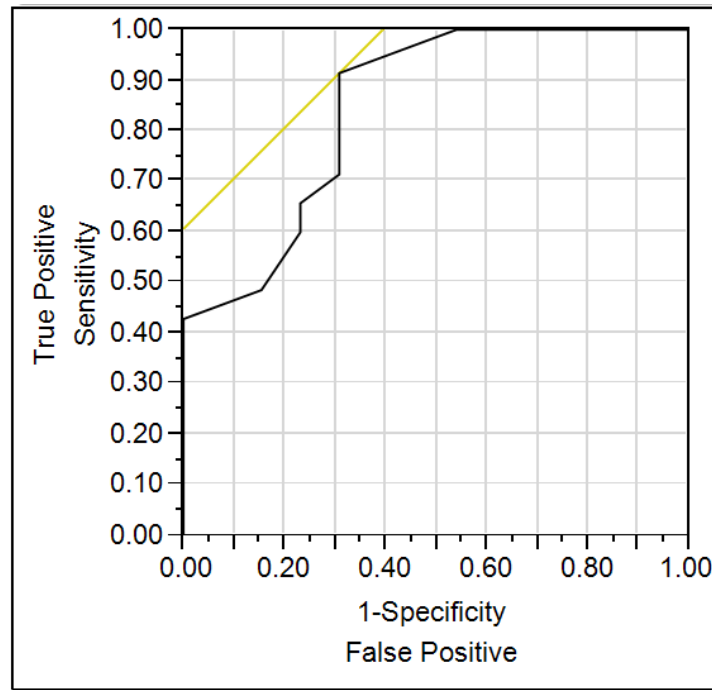


**Figure 7. ROC Curve for FFF Function Logistic Regression**

**Confusion Matrix**

|  |  | Predicted | |
|---|---|:---:|:---:|
|  |  | -1 | 1 |
| Actual | -1 | 9 | 4 |
|  | 1 | 3 | 32 |

**Figure 8. FFF Function Logistic Regression Confusion Matrix**

## 3.4 Validation Methods

The validation analysis utilized the data within the FF1 panel matrix most representational of the design space modeled by Peyton (2012) and Koslow (2012). The

target for these validation runs was a representative aircraft composite.  The defined conditions indicating sufficient data for comparison set forth by Peyton and Koslow were at least three time steps of flash data associated with a shot.  Of the original 109 FF1 runs initially deemed representative of the modeled space, only 11 contained enough data from a resultant flash to allow for comparison.

Twenty four additional runs were executed within the actual design space separately from the FF1 tests for further comparison.  These additional 24 runs used the same aluminum target (2024 T3XX) considered by Peyton and Koslow when building their models.  This augmented validation utilized a $2_{IV}^4$ factorial design with eight repeated points.  Table 8 shows the factor settings for these additional runs and the standard order.  Factor levels with two test numbers indicate the 8 replicated runs.  Of these 24, 23 produced flash data adequate for validation of the front face Peyton model, and 20 produced adequate data for validating the back face Koslow model.  These runs were designed to challenge the models in the center of their design space.

**Table 8. Flash Model Verification Augmented Runs**

## Inner Space Validation Runs

| Test # | STD | Run | Velocity (fps) | Obliquity Angle (deg) | Thickness (in) | Frag Size (grains) |
|--------|-----|-----|----------------|-----------------------|----------------|--------------------|
| 282 | 4 | 1 | 5000 | 15 | 0.25 | 40 |
| 283 & 302 | 0 | 2 | 5000 | 15 | 0.16 | 40 |
| 284 & 297 | 10 | 3 | 5000 | 30 | 0.16 | 75 |
| 285 | 7 | 4 | 7000 | 30 | 0.25 | 40 |
| 286 | 11 | 5 | 7000 | 30 | 0.16 | 75 |
| 287 & 298 | 5 | 6 | 7000 | 15 | 0.25 | 40 |
| 288 | 2 | 7 | 5000 | 30 | 0.16 | 40 |
| 289 & 300 | 12 | 8 | 5000 | 15 | 0.25 | 75 |
| 290 | 1 | 9 | 7000 | 15 | 0.16 | 40 |
| 291 | 13 | 10 | 7000 | 15 | 0.25 | 75 |
| 292 | 14 | 11 | 5000 | 30 | 0.25 | 75 |
| 293 & 305 | 9 | 12 | 7000 | 15 | 0.16 | 75 |
| 294 & 304 | 15 | 13 | 7000 | 30 | 0.25 | 75 |
| 295 & 303 | 3 | 14 | 7000 | 30 | 0.16 | 40 |
| 296 & 301 | 6 | 15 | 5000 | 30 | 0.25 | 40 |
| 299 | 8 | 16 | 5000 | 15 | 0.16 | 75 |

The detailed methodology of capturing the responses is found in Bestard and Kocher (2010). Peyton (2012) and Koslow (2012) provided the flash models. These models predict a time-series output that was compared to the actual time based flash response from live fire test shots. The validation technique used is illustrated in both Peyton (2012) and Koslow (2012), but is reviewed quickly here for continuity purposes.

The coefficients for the front and back face models were stored in a dataset within MatLab. A table of these coefficients for both front and back face is found in Appendix B. The predicted flash radii were calculated using these coefficients and the corresponding factor settings of the specific runs which produced sufficient post processed flash data. The average flash radius and the cumulative radii for each run was

output.   The post processed data was then used to calculate the average flash radius and cumulative radii for each run.  The graph of the post processed data values (actual) and the graph of the model's predicted values were plotted as well as the difference between the two. These graphs allowed for a visual comparison of the actual and predicted values. Runs from the FF1 matrix which lay outside the modeled space due to material type were predicted with both coefficients from the 2024 and 7075 material model settings to investigate which modeled material would best represent the honeycomb composite used as the target

The additional 24 augmentation runs were validated using a direct comparison of the actual and predicted radius values at each time step.  Some of the augmented test runs showed a "white-out" across the frame for several time steps at the beginning of the run. This "white-out" was caused by the initial flash being so large it filled the camera frame. Increasing the distance between the camera and the target may reduce this trend but may also decrease the ability to capture further flash details.  An actual distance was not provided by the range.

# 4. Analysis and Results

All six of the measure responses, front face flash, back face flash, panel weight change, penetration, front face function, and back face function from the 7.62 API comparison test were analyzed although only front face and back face flash, and penetration probability were defined by the 46th Test Group as responses pertinent to the objectives for the comparison live fire test. The additional responses were analyzed to help verify and validate the additional objectives of this research which included the capture of design and validation methodologies.

## 4.1    Analysis

### 4.1.2   Continuous Response ANOVA

The analysis for all continuous response variables was done using Design Expert from Statease and JMP9 from SAS. This data was collected and maintained by InDyne and reduced by Skyward Ltd.

### 4.1.3   Dichotomous Variable Regression

All analysis for the three dichotomous (yes/no) responses used JMP 9 software's Logistical Regression tool and an Excel spreadsheet utilizing the raw formulas for the probability of y ($\pi(x)$) and the likelihood function.  Only the front-face flash data could be analyzed because of instability in the regression model. This instability was caused by uniform responses with a constant level for one factor across multiple levels of the other factors. A constant response forces the logistical regression iterative process of fitting an S-curve to expand towards negative infinity or infinity when calculating the coefficient

values. JMP's software ends the iteration by default but the resulting model is unstable and inconclusive. Reviewing the response values in a three-dimensional table, as shown in Table 9 and Table 10, reveals the input factors whose intervals may be too large or small to result in a change of the response. These factor levels may be reconsidered for future testing if it is desired to find a response region with a probability greater than zero but less than one.

**Table 9. 3-D Back Face Flash Function**

**3 Dimensional BFF Function Response Variance**

| | | Obliquity | | | |
|---|---|---|---|---|---|
| | | 0 | | 60 | |
| | | Velocity | | Velocity | |
| | | 1500 | 2500 | 1500 | 2500 |
| Threat | -1 | 0.666667 | 0.5 | 0 | 1 |
| | 1 | 1 | 0.5 | 0.666667 | 1 |

**Table 10. 3-D Projectile Penetration Function**

**3 Dimensional Penetrate Function Response Variance**

| | | Obliquity | | | |
|---|---|---|---|---|---|
| | | 0 | | 60 | |
| | | Velocity | | Velocity | |
| | | 1500 | 2500 | 1500 | 2500 |
| Threat | -1 | 1 | 1 | 0 | 1 |
| | 1 | 1 | 1 | 0.166667 | 1 |

## 4.2 Results

### 4.2.1 Front Face Flash Duration

ANOVA indicated that specific factors were significant in describing the variance of front face flash duration (see Table 11). All factors were analyzed as nominal scale within the coded region. The model was analyzed for any normality assumption violation. It was noted that the residual of the model did not fit a normal distribution as shown in Figure 9 having a Shapiro-Wilk W test of 0.864324 with a corresponding p-value of <0.0001. This shows the residuals are not from a normal distribution. There was no apparent violation of the residuals having constant variance as evidenced by Figure 10.

**Table 11. FFF Duration ANOVA from JMP9**

### ANOVA FFF Duration

| Source | DF | Sum of Squares | Mean Square | F Ratio | Prob > F | |
|---|---|---|---|---|---|---|
| Model | 10 | 2.07E+09 | 207025166 | 9.9137 | <.0001 | significant |
| Threat | 1 | 1.077E+09 | 1077295994 | 51.588 | <.0001 | significant |
| Obliquity | 1 | 985562.74 | 985562.736 | 0.0472 | 0.8292 | |
| Thickness | 1 | 75102411 | 75102410.8 | 3.5964 | 0.0657 | |
| Velocity | 1 | 273899080 | 273899080 | 13.1161 | 0.0009 | significant |
| Threat*Obliquity | 1 | 105350670 | 105350670 | 5.0449 | 0.0308 | significant |
| Threat*Thickness | 1 | 74220829 | 74220829.2 | 3.5542 | 0.0673 | |
| Obliquity*Thickness | 1 | 63849470 | 63849470.4 | 3.0575 | 0.0887 | |
| Threat*Velocity | 1 | 105035740 | 105035740 | 5.0298 | 0.031 | significant |
| Obliquity*Velocity | 1 | 142094176 | 142094176 | 6.8044 | 0.013 | significant |
| Thickness*Velocity | 1 | 92563185 | 92563184.7 | 4.4325 | 0.0421 | significant |
| Error | 37 | 772658742 | 20882668.7 | | | |
| Lack Of Fit | 5 | 523649495 | 104729899 | 13.4588 | <.0001 | significant |
| Pure Error | 32 | 249009247 | 7781538.97 | | | |
| Total Error | 37 | 772658742 | | | | |
| C. Total | 47 | 2.843E+09 | | | | |
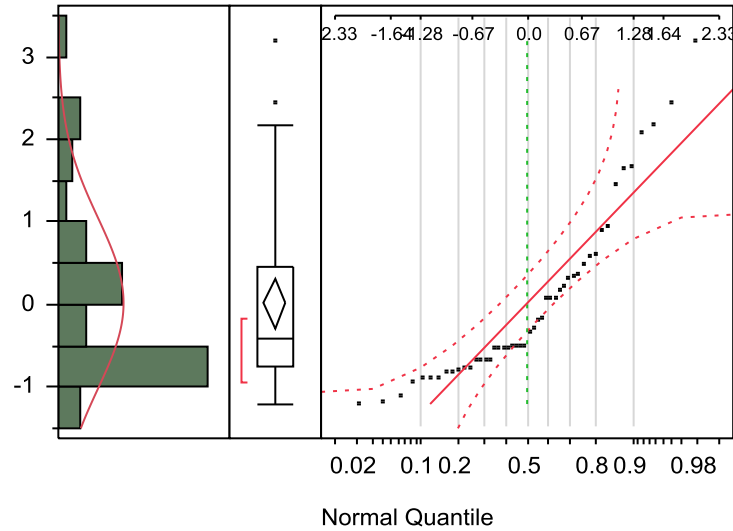
**Figure 9. FFF Duration Residual Normal Plot**
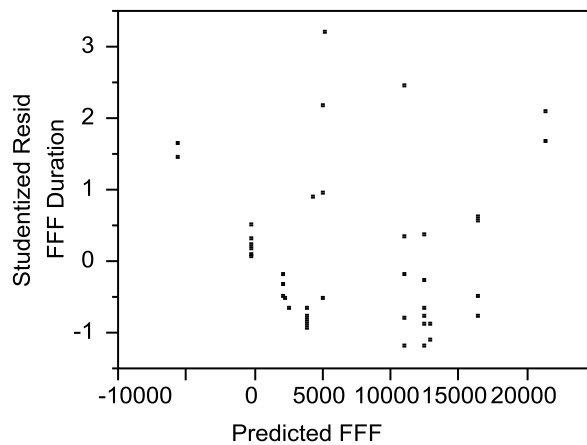


**Figure 10. FFF Duration Residuals vs. Predicted**

The noise within the system was extremely high as the ANOVA and subsequent analysis showed. This was expected due to the complex physical properties underlying the test execution. Since the primary object of this research was to determine if a difference existed between the two levels of threat used in the 7.62 comparison testing, the results are found very distinct. It was found that there is a statistically significant difference between the two threat levels with respect to the response of front face flash.

44

Threat level "1" (54mm) causes a statistically higher FFF duration as seen in Table 12. Front Face Flash Duration.

**Table 12. Front Face Flash Duration**

## FFF Duration

|          | 39mm     | 54mm      |
| -------- | -------- | --------- |
| **Min**      | 0        | 0         |
| **Max**      | 13775    | 29340     |
| **Median**   | 305.5    | 9986.5    |
| **Average**  | 1667.542 | 11551.667 |

### 4.2.2 Back Face Flash (BFF) Duration

As with FFF duration the ANOVA indicated certain factors were significant in describing the variance of back face flash duration. The resulting ANOVA is shown in Table 13. All factors were set to nominal within the coded region. Assumption violations were noted with residuals of the model not fitting a normal distribution. This is shown in Figure 11 having a Shapiro-Wilk W test of 0.912501 with a corresponding p-value of <0.0016. There was no apparent violation of the residuals having constant variance when plotted against the predicted values as shown in Figure 12.

Table 13. BFF Duration ANOVA from JMP9

## ANOVA BFF Duration

| Source | DF | Sum of Squares | Mean Square | F Ratio | Prob > F | |
|---|---|---|---|---|---|---|
| Model | 10 | 1.91E+09 | 190506791 | 9.0257 | <.0001 | significant |
| Threat | 1 | 5.33E+08 | 533069410 | 25.2553 | <.0001 | significant |
| Velocity | 1 | 341141 | 341141 | 0.0162 | 0.8995 | |
| Obliquity | 1 | 1.88E+08 | 187666694 | 8.8911 | 0.005 | significant |
| Thickness | 1 | 4.78E+08 | 477739844 | 22.6339 | <.0001 | significant |
| Threat*Velocity | 1 | 1437547 | 1437547 | 0.0681 | 0.7956 | |
| Threat*Obliquity | 1 | 2.09E+08 | 208946264 | 9.8993 | 0.0033 | significant |
| Velocity*Obliquity | 1 | 1.06E+08 | 105945109 | 5.0194 | 0.0312 | |
| Threat*Thickness | 1 | 4.32E+08 | 431598631 | 20.4479 | <.0001 | significant |
| Velocity*Thickness | 1 | 3187782 | 3187782 | 0.151 | 0.6998 | |
| Obliquity*Thickness | 1 | 2.78E+08 | 277753807 | 13.1592 | 0.0009 | significant |
| Error | 37 | 7.81E+08 | 21107236 | | | |
| Lack Of Fit | 5 | 6.31E+08 | 126137199 | 26.8588 | <.0001 | significant |
| Pure Error | 32 | 1.5E+08 | 4696303.9 | | | |
| C. Total | 47 | 2.69E+09 | | | | |
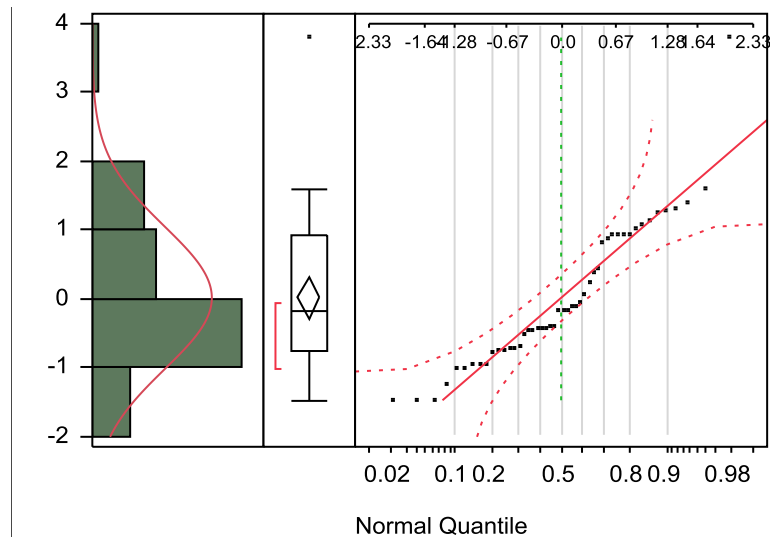


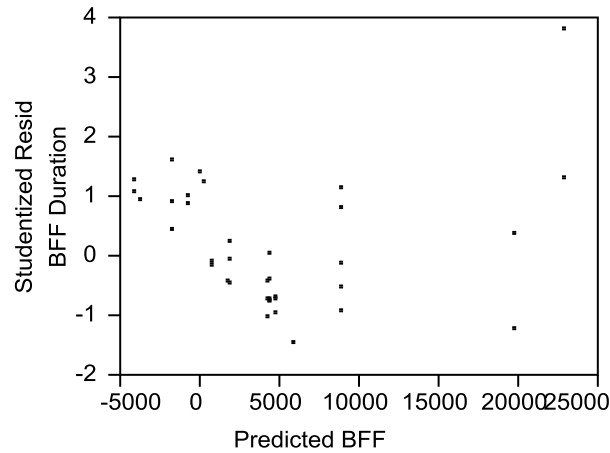Figure 11. BFF Duration Residual Normal Plot

**Figure 12. BFF Duration Residuals vs. Predicted**

The system exhibited a large amount of noise and model assumption violations do

not allow for an accurate statistical model for the response, much like the results of the

FFF duration analysis. Once again the objective of the test was to statistically infer

whether there was a difference between the two threats. The duration times for BFF are

shown in Table 14. It was found that threat level "1" (54mm), was significant higher

than BBF duration at level "-1"(39mm).

**Table 14. Back Face Flash Duration**

## BFF Duration

|          | 39mm     | 54mm     |
|----------|----------|----------|
| **Min**    | 0        | 0        |
| **Max**    | 5055     | 37468    |
| **Median** | 976      | 1916.5   |
| **Average**| 1288.125 | 6845.083 |

### 4.2.3   Panel Weight Change

The change in panel weight due to projectile impact was analyzed although it was

not an objective defined at the onset of this research. Unlike FFF and BFF duration,

change in the weight of the target panel showed no dependence upon the input factors and required no additional analysis.

### *4.2.4   Front Face Flash Function*

Front face flash (FFF) function was analyzed using logistic regression to provide additional statistical support when deciding if the results of the threat variance influence upon flash duration were of practical significance. The odds ratio was a simple calculation because the model was linear and indicated no significant interaction terms. Because of this the ratio was simply the exponential of the estimated coefficients or $e^{\beta_i}$. Table 6 and Table 7 from section 3.3.2 provide the analytical results and are reintroduced here.

**Table 6. Front Face Flash Function Parameter Estimates from Logistic Regression**

**Parameter Estimates and Significance**

| Term | Estimate | Lower 95% | Upper 95% | Std Error | Chi Square | Prob>ChiSq | |
|---|---|---|---|---|---|---|---|
| Intercept | 1.408933 | -4.208178 | 7.5483201 | 2.908917 | 0.23 | 0.6281 | |
| Threat | 1.49406 | 0.5765264 | 2.7206658 | 0.5303128 | 7.94 | 0.0048 | *significant* |
| Natural Velocity | -0.0004 | -0.002206 | 0.0013026 | 0.0008671 | 0.21 | 0.6437 | |
| Natural Obliquity | 0.04966 | -0.089929 | -0.019226 | 0.017521 | 8.03 | 0.0046 | *significant* |
| Natural Thickness | -4.23126 | -27.37042 | 15.16441 | 10.36086 | 0.17 | 0.6809 | |

**Table 7. FFF Function Odds Ratio for Unit Increase in Input Factors**

**Unit Odds Ratio**

| Term | Odds Ratio | Lower 95% | Upper 95% |
|---|---|---|---|
| Threat | 19.84835 | 3.1678489 | 230.74926 |
| Natural Velocity | 0.999599 | 0.997797 | 1.001303 |
| Natural Obliquity | 1.050913 | 0.913996 | 0.980957 |
| Natural Thickness | 0.014105 | $1.3 \times 10^{-12}$ | 3853183 |

Parameters for threat and natural obliquity showed statistical significance predicting front face flash (FFF) function. Given their relevance to the model, the calculated odds ratio was applied to response with some interpretive meaning. Interpreting this statistic, perhaps in a more understandable way, it can be stated that given the probability of a function at an obliquity of 15 degrees, changing the obliquity to 16 degrees increase the odds of a function by a factor of 1.050913. Given the probability of a function at threat level -1 (39mm) changing to threat level 1 (54mm) increases the odds of a function by a factor of 19.85. It is interesting to note that over the entire range of the obliquity the odds ratio is approximately equal to the odds ratio for change in threat level.

Figure 13 shows the front face flash function odds ratio for both levels of threat across the obliquity range. The difference between the two threats becomes less in the odds space as obliquity increases but the difficulty or the effort of bridging this gap remains the same across the entire range.
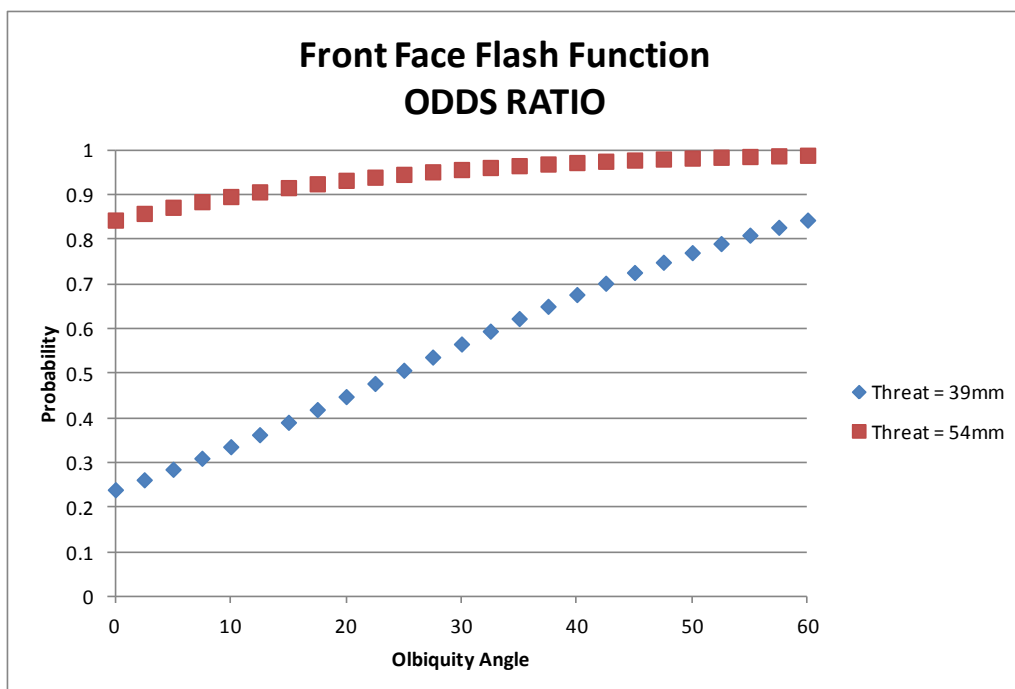
**Figure 13: FFF Function Odds Ratio**

# 5. Flash Function Model Validation

Validation of the Peyton (2012) and Koslow (2012) models involved the methodology outlined in Section 3.4.  Flash results were compared visually and numerically to the predictions from the Peyton and Koslow model.   Runs from the FF1 test matrix were used to initially validate the model.  This was done to evaluate the models robustness across a different material from which the model was derived.  The models were constructed based on an aluminum target of varying thicknesses and the FF1 runs used in the comparison of this model targeted a honeycomb composite.

The complete set of plots and the numerical values derived by the validation methodology for the aircraft composite (FFI runs) can be obtained through AFIT/ENS. Plots for the aluminum targets (24 augmented runs) are in Appendix C.   The aircraft composite flash results were compared to predicted model values using coefficients for both the 2024 and 7075 materials.  A sample of the FFI comparison of the front face flash using the 2024 material model coefficients is shown in Figure 14, the 7075 material coefficients in Figure 15.  Plots for the back face flash prediction versus actual showed the same results.  All plots for FF1 validation runs showed the models tended to over-predict flash radius in both dimensions.  These validation results showed that the models were not adequate for runs outside the design space specifically on materials beyond which the models were built.
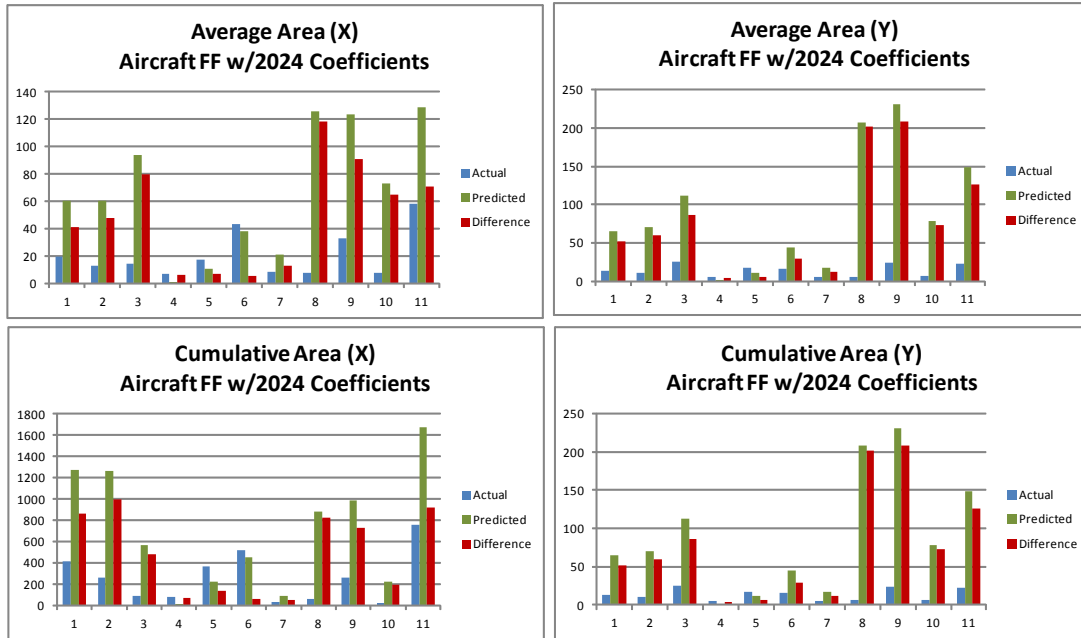
**Figure 14. FF1 Validation Using 2024 Coefficients**
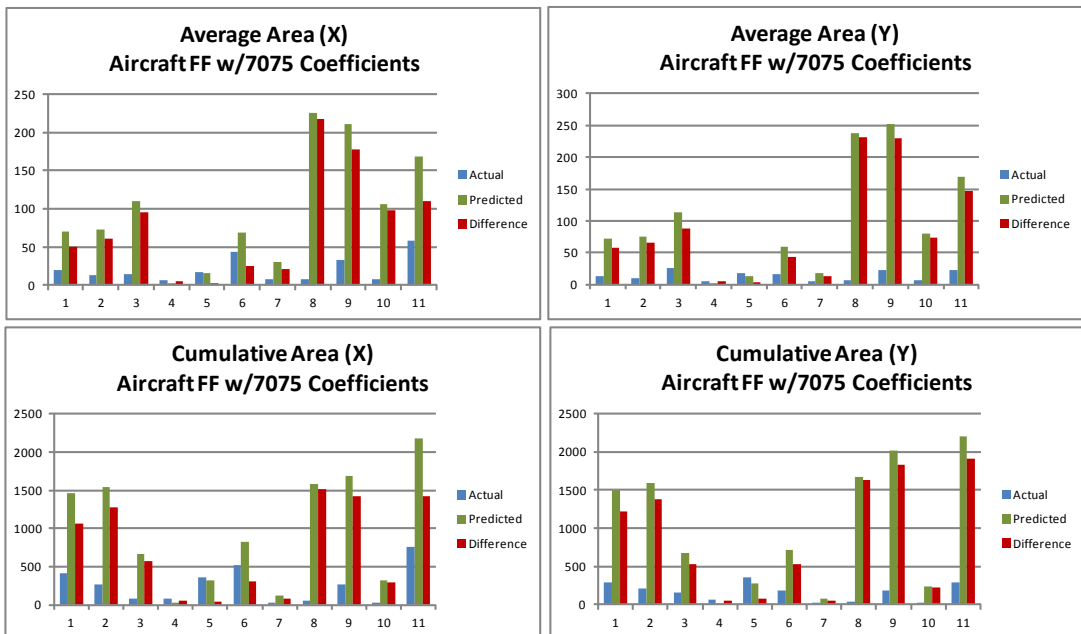


**Figure 15. FF1 Validation Using 7075 Coefficients**

The validation of the model within the center of its design space utilizing an

aluminum target showed the opposite effect, the model tended to under predict. Figure

16 shows the average and cumulative comparison for the front face flash of the

augmented validation runs against a 2024 aluminum target. Figure 22 shows the same tendency existed for the back face flash. These runs were all conducted in the center of the space which the model was built to describe.
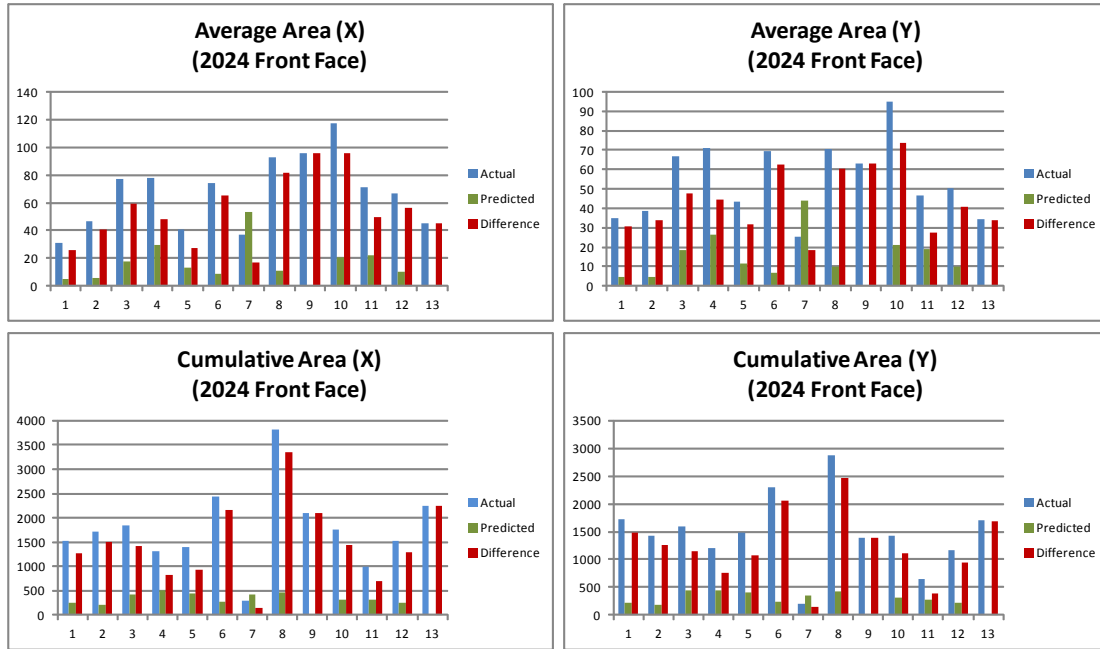


**Figure 16. Front Face Augmented Validation**
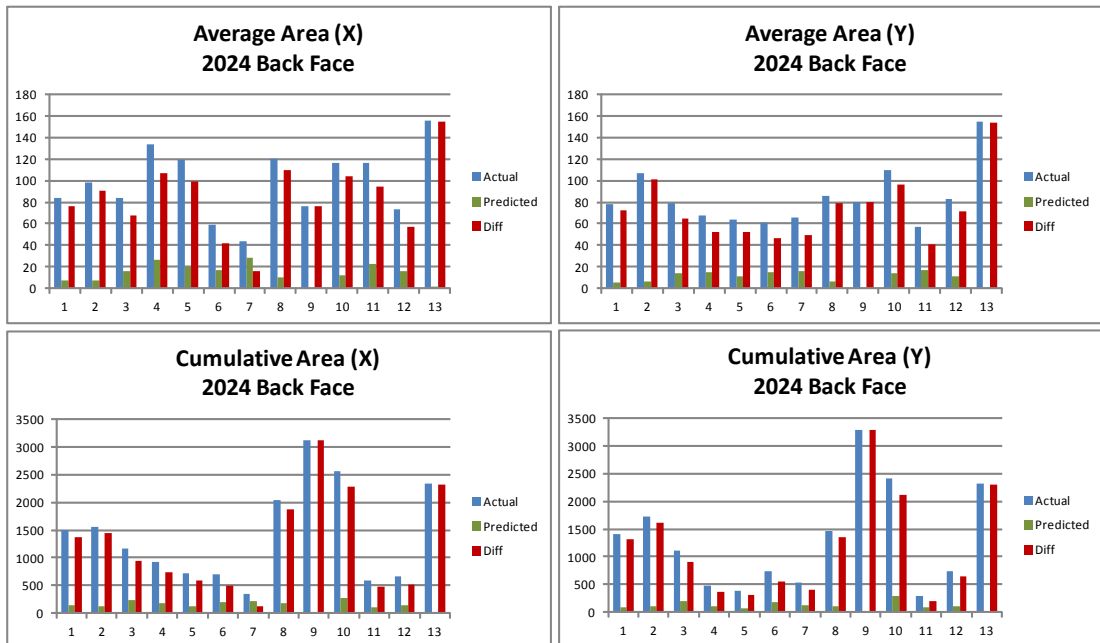


**Figure 17. Back Face Augmented Validation**

To further clarify the finding of model inadequacy indicated by the average comparisons, the predicted radius was compared against the actual radius for individual time steps on each of the 23 aluminum target runs which produced sufficient flash data. Figure 18 shows a sample of the resulting plot. Appendix C contains the full suite of graphics depicting the test shot data versus model predictions. Whether front-face or back-face, the model tends to under predict the flash for target material within the model space, particularly early in the shot event. Such bias was noted by both Peyton (2012) and Koslow (2012). Thus, this research cannot say the current flash model is doing an adequate job of predicting flash. To overcome the initial under-estimation bias, model developers may want to consider some form of weighted curve fitting.
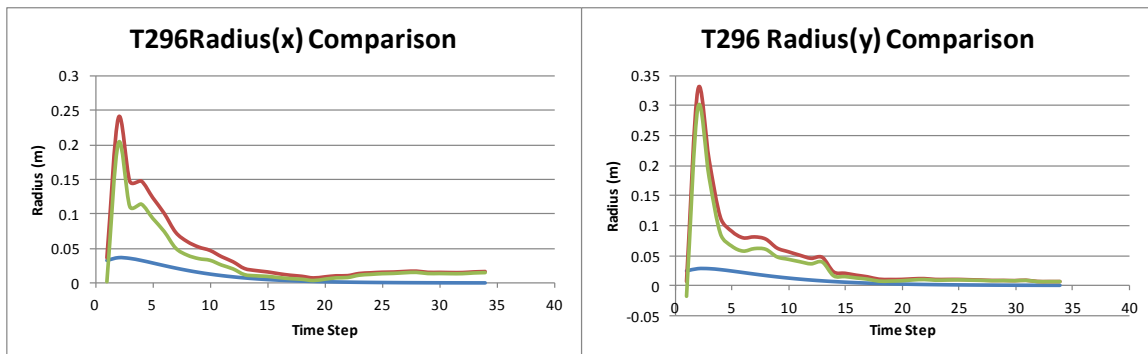


**Figure 18. Time Step Comparison of Flash Radius**

It is apparent that while the results from the FF1 validation could be excused due to the factors being outside the space for which the model was designed, the augmented validation runs within the center of the actual design space shows the models are inadequate.

## 6. Conclusions

The original objective at the beginning of this work was to design a test and capture the methodology used to provide a statistical inference regarding the different effects between two armor piercing incendiaries (API). Unified oversight on the design process from beginning, through execution, to final analysis is critical to capturing the maximum amount of information that can be provided from appropriately planned testing. Of particular note, the correct analysis method must be used to statistically validate the designed model's objectives. Live fire testing is particularly susceptible to such deviations because of the nature of its process; with multiple agencies involved in the design, execution, analysis and implementation of test results. Volatility of customer requests, administrative oversight and resource availability, all impact not only the live fire design and execution but the factors that drive the reasoning behind the initial design choices.

Past testing has utilized the capabilities of statistical rigorous designs but have lacked severely in the documentation of such efforts. At times this has been attributed to the sensitive nature of many of the projects that are executed within the DoD. This sensitivity often restricts even the program definition of the factors and their levels used in the testing. This work argues that while the details of the programs being classified as too sensitive for general publication, the statistical methods used to build and analyze these tests should not fall within this classification restriction. The methodology behind the design and analysis can be provided without disclosure of any information determined to be unacceptable for public release. Any methodology used in a classified test came from the public arena and is itself not indicative of the sensitivity level of the test. With

over $75 billion spent by DoD on testing each year (Johnson, Hutto, Simpson, & Montgomery, 2012), the discipline of statistical testing would benefit immensely from the shared methodologies of test design within the DoD.

It was determined by this research that there is a statistical difference between API projectiles in regard to the responses of front face flash, back face flash, and probability of a flash function. A 54mm API increases flash duration and the odds of a function occurrence when compared to a 39mm API. Validation of the Peyton and Koslow models indicate the current models are inadequate at predicting the radius of a flash function caused by a ballistic impact.

## 6.1    Recommendations

The use of statistical rigor in test design and execution has increased interest within the DoD. With this increased emphasis many programs have implemented multiple requirements to ensure that statistical insight and inference methods are examined when building the test plan for a system. Large systems in particular are under heavy scrutiny to use these methods to reduce cost and deliver more with less. However, such emphasis has also increased the use of statistical rigor in word only. As found by this research, often those charged with enforcing statistical requirements are unfamiliar with the methods these requirements enforce. A simple statement within a test plan indicating that statistically rigorous methods were examined is often considered sufficient to fulfill the requirement. Familiarity with the theoretic background and implementation of these methods needs to be stressed for both those executing the design and analysis of the tests and the agencies creating these requirements of statistical rigor. Even

organizations with statistically certified and experienced personnel fall short of reaching

the full potential statistical methods can provide due to the gap between the theoretical

development of the design and the execution/collection of the data generated by the

testing.  Tests often span across three or more agencies from the initial discussions

driving the selection of test factors and objectives to the final analysis and written report

detailing the outcome of the test.  Often the continuity behind the reasoning and

motivations of the test are lost as the procedure progresses.  Within the DoD in particular

personnel volatility creates gaps within the testing continuity as large test programs span

over multiple years.  More times than not numerous contacts within the multiple agencies

share the responsibility of passing the knowledge for the design, data, methodology of

conduct and through this process information lost.  This information can be anything

from the reasoning behind the design choice to the levels of factors to be tested.

## 6.2    Future Research

Future work within live fire test could expand the research of capturing

methodologies currently used across differing agencies to produce a single methodology

to be utilized by all agencies and departments within the DoD.  The attempt to capture the

varying and wide demands within the live fire test discipline may help to reduce the

frequent deviations from the plans established during a certified statistically rigorous test

plan.  Familiarizing the operators and data collectors of the systems and programs under

test could bridge the knowledge gap described by Coleman and Montgomery and which

this research confirmed to exist within even the best planned tests.  Particularly in DoD

programs the volatility of personnel introduces challenges not understood in the private sector.

Future research could include further investigation within the Peyton and Koslow model to adjust for prediction of function shape and duration across a production representative composite material as well as the generation of a function characterization model for armor piercing incendiary projectiles. Work within these models could incorporate a coefficient in the initial regression to account for the expansion of the model space to include all variants of aluminum and composite materials currently being used in the production or design of aircraft in the nation's inventory. Future work may also be used to explore the reasoning behind the discrepancy of the model within the center space as found during the augmented validation against aluminum targets. Multiple options of doing this may include using a combined data set of the augmented runs of this work and the data used in the regression by Peyton and Koslow to estimate the model coefficients as well as a weighted regression to overcome initial under-estimation by the models. Future work could also re-examine the biased discovered by Tallafuse (2011) and confirmed in this work regarding the white-out effect during post processing of digital video data and its correlation to test set up.

# 7.62 mm API Comparison White Paper

**Objective:** Determine if variances exist between the 7.62 x 39mm API (Type BZ) and the 7.62 x 54mm API (Type B-32) in terms of penetration and function characteristics against 2024-T3 Aluminum

**Hypotheses:**   Null: $H_0: \sigma_1^2 \ = \ \sigma_2^2$ (Variances are equal)
Alternative: $H_1: \sigma_1^2 \ \neq \ \sigma_2^2$ (Variances are unequal)

**Assumptions:** Independence, Distributed Normally, Equal Variance (F-Statistic – Two Tail)

**Significance Level:** Test at 95% ($\alpha = 0.05$) and 99% ($\alpha = 0.01$)

**Control Factors:**
- Threats (7.62 x 39mm API Type BZ and the 7.62 x 54mm API Type B-32) – Categoric
- Velocities (Low and High) – Numeric (Continuous)
- Obliquities (0° and 45°) – Numeric (Nominal)
- Thickness (0.125" and 0.25") – Numeric (Nominal)

**Response Variables:**
- Function Duration
- Function Size (i.e., Maximum Area)
- Function Location (i.e., front or back of panel)
- Function Maximum Distance from Impact Point
- Residual Velocity/Impact Velocity Ratio (i.e., Percentage)
- Residual Mass/Impact Mass Ratio (i.e., Percentage)
  - Percentage mass loss may be misleading…pretest, the mass is of the whole projectile (core, incendiary, jackets, etc.), but post-test the mass is only of the core.
  - The threat types have different incendiary amounts (2 gr vs. 10 or 15 gr) so not sure percentage is the right metric
- Anything Else? (Hole size / area removed ect.)

**Nuisance Factors:**
- Controllable
  - Threat
    - Threat loading (Which RSO loads the threat, which RSO loads the threat in the gun, gun powder amt, how powder packed, how powder weighed, how many times each threat casing is used etc.)
    - Incendiary material variance (Will testing be done on this in time to use?)
  - Test Setup
    - How many time gun fired before cleaned

- How many times gun is fired per day (first shot of the day on a cold barrel versus last shot of the day on a warm barrel) Can temp of barrel be measured?
- Number of time each panel is impacted (if more than once)
- What technicians load panels, setup cameras, setup instrumentation, measure weights of panels, calculates impact/residual velocities, measures pre-impact/residual weights of threats, etc.
- Uncontrollable
  - Threat
    - Variation of incendiary amount per threat type and per physical threat (i.e., within the same threat type)
    - Threat design (i.e., core mass, core length, core material, etc.)
    - Threat lot number
  - Test Setup
    - Environmental conditions of range (Temp, humidity, etc.)
    - How velocity was measured (error inherent in measurement type, who set up gun breaks and events and how they did it, etc.)
    - How many shots have been conducted over the life of the barrel
    - Variability in gun distance
    - Variability in gun angle (should be consistent, but gun will move between each shot)
    - Target material lot (map shot panels from raw material sheet)
    - How threat impacts target panel (yaw/pitch/roll angles of threat impact)
- Anything Else?

## Potential Designs:
- $2^4$ Full Factorial using Design Expert (3 reps - 48 runs)
- $2^4$ Full Factorial using JMP (3 reps - 48 runs)
- $2^{4-1}$ Fractional Factorial using Design Expert (3 reps - 24 runs)
- Custom Factorial using JMP (3 reps - 52 runs)

## Notes:
- Completely randomizing the design is not feasible in terms of cost and range time
- Suggests using blocking techniques to complete the design in blocks for each the 0° and 45° obliquities
- Randomization will be conducted within each block to form a Randomized Complete Block Design (RCBD)
- With 200+ inventory, perhaps conduct smaller exploratory experiments to extract significant factors and use remaining runs to get as much data on those as possible.
- Possibly additional factors could be input from 46th.

*Appendix B. Flash Radius Meta Model Coefficient Tables*

## Front Face Coefficients

| Radius | X | | Y | |
|---|---|---|---|---|
| Coefficient | $\beta_x$ | $\gamma_x$ | $\beta_y$ | $\gamma_y$ |
| velocity ($b_1$)** | 7.29E-06 | 0.000182 | -4.00E-05 | 0.000264 |
| obliquity ($b_2$) | 0.0126068 | 0.037188 | 0.0113538 | 0.038654 |
| thickness ($b_4$) | 0.6632973 | 1.348181 | 0.269234 | 1.967688 |
| mass ($b_5$)** | 0.0006022 | 0.005674 | 0.0004821 | 0.005503 |
| intercept ($b_0$) | 1.261354 | 3.531991 | 1.770376 | 2.802686 |

**measured values used for velocity and mass

## Back Face Coefficients

| Radius | X | | Y | |
|---|---|---|---|---|
| Coefficient | $\beta_x$ | $\gamma_x$ | $\beta_y$ | $\gamma_y$ |
| velocity ($b_1$)** | 1.41E-04 | 0.000444 | 6.19E-05 | 0.000738 |
| obliquity ($b_2$) | -0.00435 | 0.014743 | 0.00267 | 0.001499 |
| thickness ($b_4$) | 0.40562 | -0.36829 | 0.456721 | -3.2996 |
| mass ($b_5$)** | 0.001821 | 0.01093 | 0.002458 | 0.010505 |
| intercept ($b_0$) | 0.775179 | 2.09957 | 0.959158 | 1.193209 |

**measured values used for velocity and mass

## Appendix C. Flash Radius Validation Plots

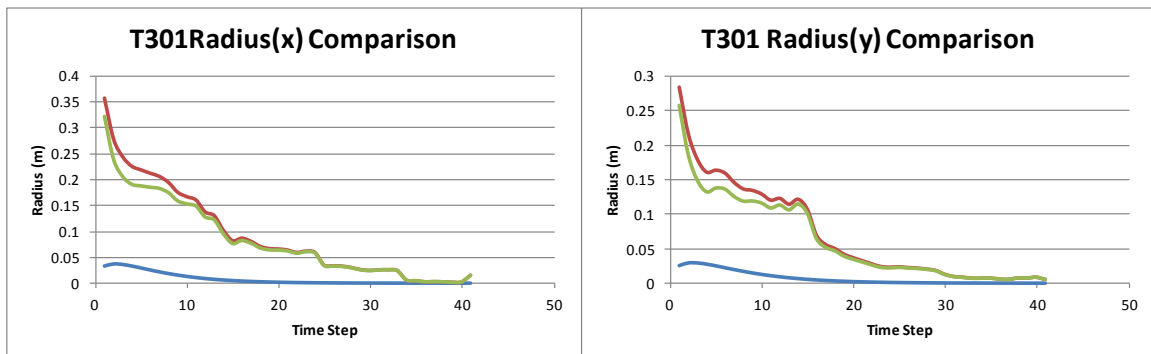BACK FACE FLASH PLOTS

Blue = Predicted               Red = Actual               Green = Difference
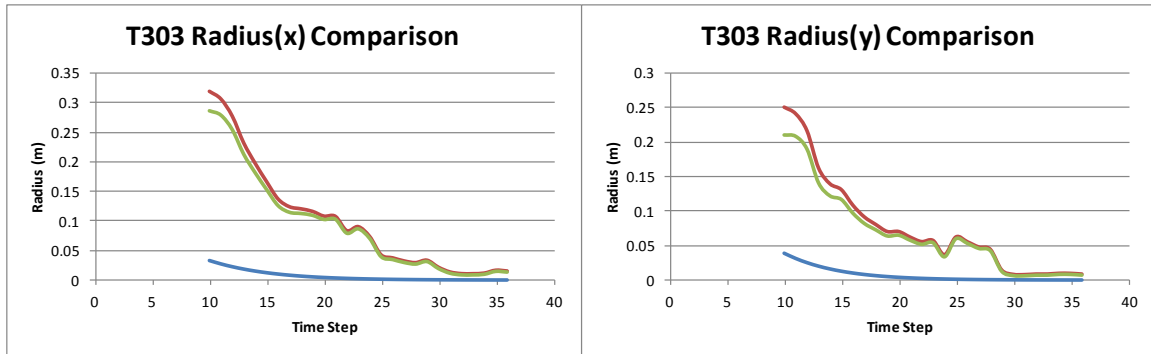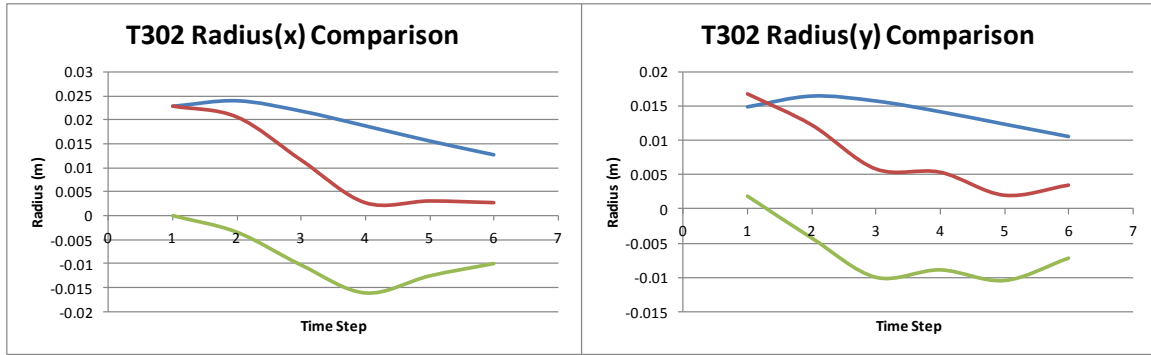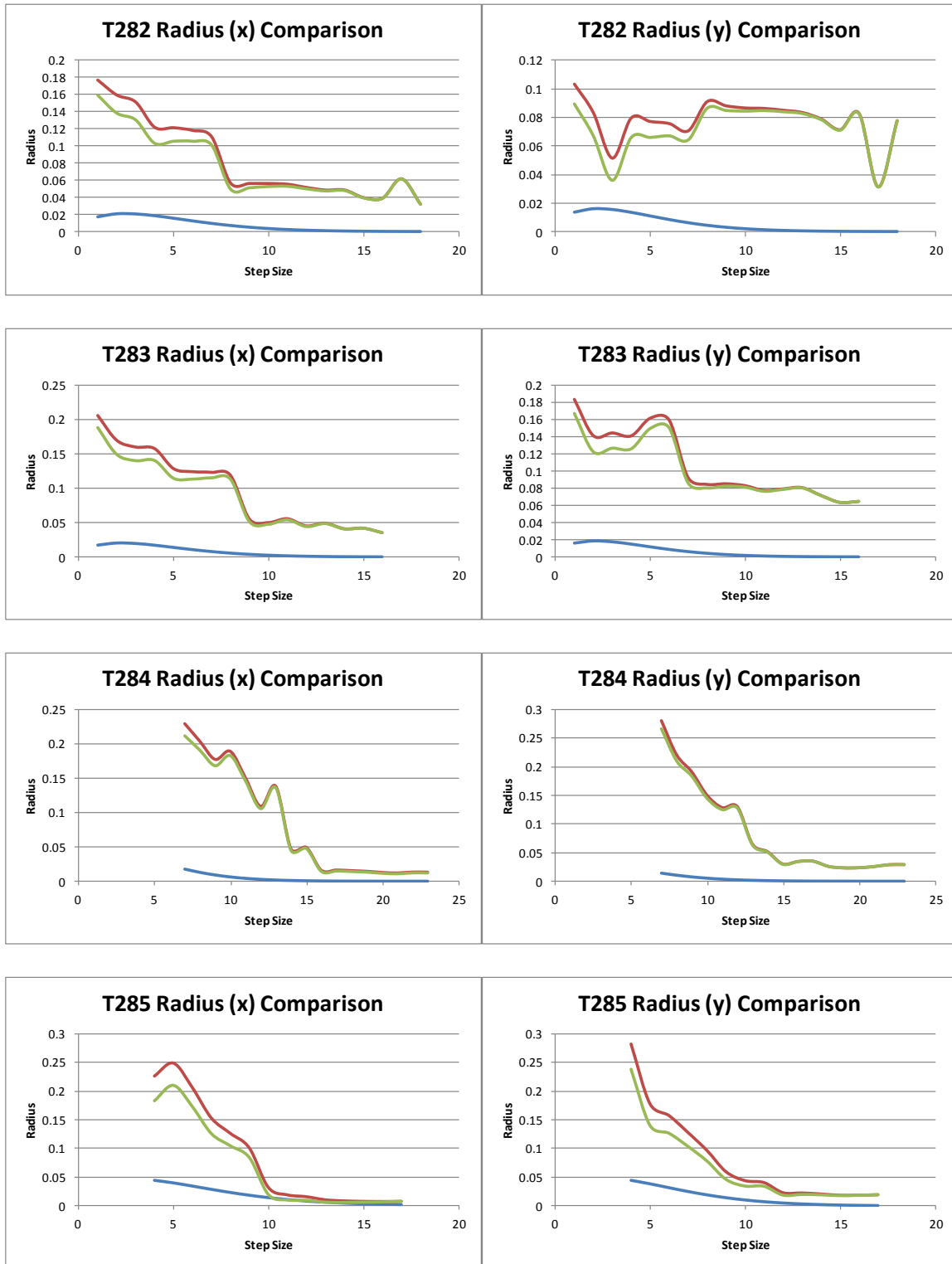
**T285 Radius(x) Comparison**

**T285 Radius(y) Comparison**

**T286Radius(x) Comparison**

**T286 Radius(y) Comparison**

**T287 Radius(x) Comparison**

**T287 Radius(y) Comparison**

**T288 Radius(x) Comparison**

**T288 Radius(y) Comparison**

**T289 Radius(x) Comparison**

**T289 Radius(y) Comparison**

**T290 Radius(x) Comparison**

**T290 Radius(y) Comparison**

**T291 Radius(x) Comparison**

**T291 Radius(y) Comparison**

**T292Radius(x) Comparison**

**T292 Radius(y) Comparison**

# T297Radius(x) Comparison

# T297 Radius(y) Comparison

# T298 Radius(x) Comparison

# T298 Radius(y) Comparison

# T300Radius(x) Comparison

# T300 Radius(y) Comparison

# T301Radius(x) Comparison

# T301 Radius(y) Comparison

## T302 Radius(x) Comparison

## T302 Radius(y) Comparison

## T303 Radius(x) Comparison

## T303 Radius(y) Comparison

## T304 Radius(x) Comparison

## T304 Radius(y) Comparison

## T305 Radius(x) Comparison

## T305 Radius(y) Comparison

67

T282 Radius (x) Comparison

T282 Radius (y) Comparison

T283 Radius (x) Comparison

T283 Radius (y) Comparison

T284 Radius (x) Comparison

T284 Radius (y) Comparison

T285 Radius (x) Comparison

T285 Radius (y) Comparison

# T286 Radius (x) Comparison

# T286 Radius (y) Comparison

# T287 Radius (x) Comparison

# T287 Radius (y) Comparison

# T288 Radius (x) Comparison

# T288 Radius (y) Comparison

# T289 Radius (x) Comparison

# T289 Radius (y) Comparison

# T290 Radius (x) Comparison

# T290 Radius (y) Comparison

# T291 Radius (x) Comparison

# T291 Radius (y) Comparison

# T292 Radius (x) Comparison

# T292 Radius (y) Comparison

# T293 Radius (x) Comparison

# T293 Radius (y) Comparison

T298 Radius (x) Comparison

T298 Radius (y) Comparison

T300 Radius (x) Comparison

T300 Radius (y) Comparison

T301 Radius (x) Comparison

T301 Radius (y) Comparison

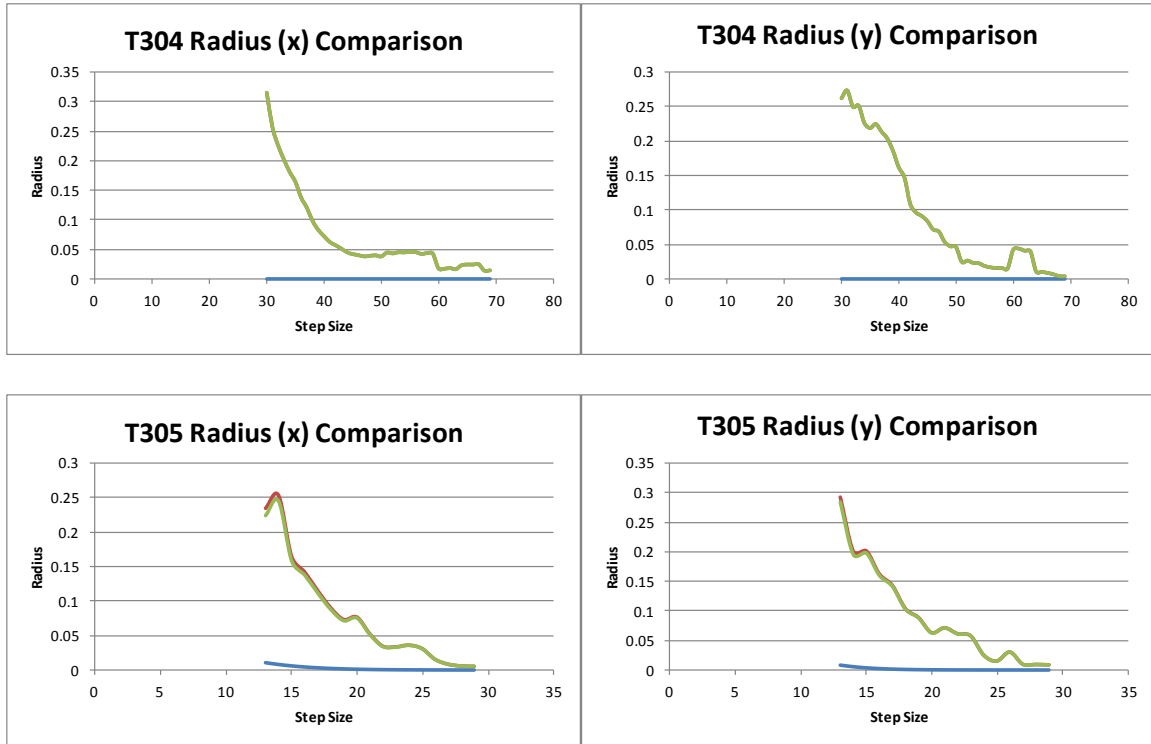T303 Radius (x) Comparison

T303 Radius (y) Comparison

### *Bibliography*

Andrus, J. (2010, Spring). JSF Live Fire Test - Pilot-in the Loop Simulator Testing. *Aircraft Survivability* , pp. 11-13.

Ball, R. E. (2003). *The Fundamentals of Aircraft Combat Survivability Analysis and Design (Second Edition).* American Institute of Aeronautics and Astronautics.

Bestard, J. J., & Kocher, B. (2010). Ballistic Impact Flash Characterization. *AIAA* , 2573.

Blythe, R. M. (1993). *Preliminary Empirical Characterization of Steel Fragment Projectile Penetration of Graphite/Epoxy Composite and Aluminum Targets.* Wright Patterson AFB, OH: MS thesis, Air Force Institute of Technology.

Cohen, A. N. (2010). *Examining Split-Plot Designs for Developmental and Operational Testing.* Wright Patterson AFB, OH: MS Thesis, Air Force Institute of Technology.

Coleman, D. E., & Montgomery, D. C. (1993). A Systematic Approach to Planning for a Designed Industrial Experiment. *TECHMOMETRICS* .

Hahn, G. (1977). Some Thing Engineers Should Know About Experimental Design. *Journal of Quality Technology* , 13-20.

Henninger, T. A. (2010). *Characterization of Ballistic Impact Flash: An Initial Investigation and Methods Development.* Wright Patterson AFB, OH: MS thesis, Air Force Institute of Technology.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied Logistic Regression.* New York: John Wiley & Sons, Inc.

Johnson, R. T., Hutto, G. T., Simpson, J. R., & Montgomery, D. C. (2012). Designed Experiments for the Defense Community. *Quality Engineering* , 60-79.

Kemp, J. S., & Woods, L. H. (2011). Live Fire Testing A Legacy Wing Assessing Dry Bay Fire Potential in the C-5 Wing. *52nd AIAA/ASME/ASCE/AHS/ACS Structures, Structual Dynamics and Materials Conference.* Denver, Colorado: American Institute of Aeronautics and Astronautics.

Knight, E. E. (1992). *Predicting Armor Piercing Incendiary Projectile Effects After Impacting Composite Material.* Wright Patterson AFB, OH: MS thesis, Air Force Institute of Technology.

Koslow, M. J. (2012). *Ballistic Flash Characterization: Penetration and Exit-Side Flash.* Wright Patterson AFB, OH: MS thesis, Air Force Institute of Technology.

Lanning, J. W. (1993). *Predicting Armor Piercing Incendiary Projectile Effects After Impacting Two Composite Panels.* Wright Patterson AFB, OH: MS thesis, Air Force Institute of Technology.

Menard, S. (2002). *Applied Logistic Regression Analysis.* Thousand Oaks: Sage Publications.

Montgomery, D. C. (2008). *Design and Analysis of Experiments 7th Edition.* John Wiley & Sons, Inc.

National Research Council. (1995). *Live Fire Testing of the F-22.* Washington, DC: The National Academies Press.

National Research Council. (1993). *Vulnerability Assessment of Aircraft: A Review of the Department of Defense Live Fire Test and Evaluation Program.* Washington, DC: The National Academies Press.

Office of the Secretary, o. D. (2010). *Guidance on the use of Design of Experiments (DOE) in Operational Test and Evaluation.* Washington, DC: DOT&E.

Peyton, D. J. (2012). *Ballistic Flash Characterization of Entry-Side Flash.* Wright Patterson AFB, OH: MS thesis, Air Force Institute of Technology.

Reynolds, J. K. (1991). *A Response Surface Model for the Incendiary Functioning Characteristics of Soviet API Projectiles Impacting Graphite Epoxy Composite Panels.* Wright Patterson AFB, OH: MS thesis, Air Force Institute of Technology.

Talafuse, T. P. (2011). *Emprical Characterization of Ballistic Impact Flash.* Wright-Patterson AFB, OH: MS thesis, Air Force Institute of Technology.

Tonnessen, L. (2011, Spring). An Historic Perspective on Live Fire Test and Evalutaion (LFT&E). *Aircraft Survivability* , pp. 8-10.

Wolf, S. E. (2012). *Modeling Small Unmanned Aerial System Mishaps Using Logistic Regression and Artificial Neural Networks.* Wright Patterson AFB, OH: MS thesis, Air Force Institute of Technology.

**Vita**


   Captain Chad Nile Chamberlain graduated from Lehi High School in Lehi, Utah. He entered the United States Air Force enlisted force on 1 August 2000 as an A1C and became a Biomedical Equipment Technician (BMET) after graduating as a distinguished graduate from the 11 month DoD BMET technical school Sheppard Air Force Base, Tx . His first permanent duty assignment was at 74th Medical Support Squadron at Wright Patterson Air Force Base, OH.  In 2004 he was accepted into the Airmen Education Commissioning Program and was later commissioned an officer in 2007, graduating from Wright State University with a Bachelors Degree in Electrical Engineering.

   His first assignment as a commissioned officer was at Edwards AFB, California, assigned to the 31st Test and Evaluation Squadron where he worked with the MQ-9 as a test engineer and analyst for the programs IOT&E, RQ-4 as lead engineer and program lead for the block 3 and the block 4 operational testing respectively and F-22 as the OT&E liaison for AFOTEC Detachment 6.

   In September 2010, he entered the Graduate School of Engineering and Management, Air Force Institute of Technology.  Upon graduation, he will be assigned to 711 OL Human Performance Wing at Fort Sam Houston, TX with his wife and two children.

# REPORT DOCUMENTATION PAGE

| 1. REPORT DATE (DD-MM-YYYY) | 2. REPORT TYPE | 3. DATES COVERED (From – To) |
|---|---|---|
| 22-03-2012 | **Master's Thesis** | Sep 2010 - Mar 2012 |

| 4. TITLE AND SUBTITLE | 5a. CONTRACT NUMBER |
|---|---|
| **Analysis Of Kc-46 Live-Fire Risk Mitigation Program Testing** | 5b. GRANT NUMBER |
| | 5c. PROGRAM ELEMENT NUMBER |

| 6. AUTHOR(S) | 5d. PROJECT NUMBER |
|---|---|
| Chamberlain, Chad Nile Captain, USAF | 5e. TASK NUMBER |
| | 5f. WORK UNIT NUMBER |

| 7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) | 8. PERFORMING ORGANIZATION REPORT NUMBER |
|---|---|
| Air Force Institute of Technology<br>Graduate School of Engineering and Management (AFIT/EN)<br>2950 Hobson Street, Building 642<br>WPAFB OH 45433-7765 | AFIT/OR-MS/ENS/12-06 |

| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | 10. SPONSOR/MONITOR'S ACRONYM(S) |
|---|---|
| Office of the Secretary of Defense<br>Attn: Dr. Catherine Warner<br>1700 Defense Pentagon commercial# (703) 697-7247<br>Washington D.C. 20301 e-mail: Catherine.warner@osd.mil | **OSD** |
| | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) |

**12. DISTRIBUTION/AVAILABILITY STATEMENT**
DISTRIBUTION STATEMENT A; APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

Increased emphasis from the Director of Operational Test and Evaluation (DOT&E) over the past few years to include statistical rigor in all testing has brought an augmented look at testing across the Department of Defense. This work looks at the methodology currently used in live fire testing, particularly involving risk mitigation of the KC-46 dry-bay test program. It addresses gaps within the methodology as well as designs and analyzes the results of a statistically rigorous test. In addition this research furthers recent work of modeling the characterization of ballistic impact flash by validating concurrent models and characterizing the error prone to these models as a function of time and input factors in an attempt to identify systemic bias that may be correctable.

**15. SUBJECT TERMS**
Ballistic Flash Characterization, Design of Experiments, Validation, Live Fire Testing, Statistical Rigor

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | Raymond R Hill Jr., Ph.D. (ENS) |
| U | U | U | UU | 89 | 19b. TELEPHONE NUMBER (Include area code)<br>(937) 255-3636, ext 7469; e-mail: raymond.hill@afit.edu |